DOCUMENT RESUME

ED 129 064 FL 007 928

AUTHOR	Lamberts, Frances
TITLE	Document Characteristics and Typographic Requirements in the Language Sciences. LINCS Project Document Series.
INSTITUTION	Center for Applied Linguistics, Washington, D.C. Language Information Network and Clearinghouse System.
SPONS AGENCY	National Science Foundation, Washington, D.C.
REPORT NO	CALLINCS-69-17
PUB DATE	Nov 71
GRANT	NSF-GN-771
NOTE	98p.; Some tables may be marginally legible
EDRS PRICE	MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS	Bibliographic Citations; Bibliographies; Foreign Language Periodicals; Graphic Arts; Information Dissemination; *Information Networks; Information Science; Information Services; *Information Sources; *Information Systems; *Language Research; *Linguistics; *Periodicals; Research Reviews (Publications)
IDENTIFIERS	Character Set; *Language Sciences; Transliteration; Typography

ABSTRACT

A determination of the bibliographic and graphic characteristics of current research literature in the language sciences is essential to the design of an information system intended to offer primary and secondary publication services. The results of a comparative study of bibliographic practices in language sciences secondary publications are given in the first section. The second section discusses the general typographic aspects of primary journals in the language sciences. The requirements of the character set are considered in the third section. The final section describes some of the steps taken by the pilot project, Language Information Network and Clearinghouse System (LINCS), in the experimental development of an automated bibliographic publication service. The work contains a substantial number of tables and figures. (CFM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED129064

CENTER FOR APPLIED LINGUISTICS

LANGUAGE INFORMATION NETWORK AND CLEARINGHOUSE SYSTEM (LINCS)

DOCUMENT CHARACTERISTICS AND TYPOGRAPHIC REQUIREMENTS IN THE LANGUAGE SCIENCES

By

Frances Lamberta

with the assistance of

Judith Krone

FL007928

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

LINCS PROJECT DOCUMENT SERIES / NATIONAL SCIENCE FOUNDATION GRANT
CALLINCS-69-17 November 1971 GN-771

CENTER FOR APPLIED LINGUISTICS, 1717 MASSACHUSETTS AVENUE, N.W., WASHINGTON, D.C. 20036

DOCUMENT CHARACTERISTICS AND TYPOGRAPHIC REQUIREMENTS IN THE LANGUAGE SCIENCES

By

Frances Lamberts

with the assistance of

Judith Krone

TABLE OF CONTENTS

Acknowledgments v

1. Introduction 1
2. Comparison of Bibliographic Sources 2
   - 2.1. Objectives and Background 2
   - 2.2. Sources 4
   - 2.3. Method 4
   - 2.4. Results and Comments 9
     - 2.4.1. Source Data 9
     - 2.4.2. Document-Category Data 20
   - 2.5. Summary and Conclusion 32

   References 34
3. Some Typographic Aspects of the Primary Literature 35
   - 3.1. Purpose 35
   - 3.2. Method 35
     - 3.2.1. Sample Selection 35
     - 3.2.2. Classification of Non-Textual Material 37
     - 3.2.3. Determining the Size of an Item 37
     - 3.2.4. Symbol Count 38
   - 3.3. Results 38

   References 43
4. Requirements of the Character Set: A Sample of Literature in Computational Linguistics 44
   - 4.1. Purpose 44
   - 4.2. Procedure 44
     - 4.2.1. Selection of Sample 44
     - 4.2.2. Method 47
   - 4.3. Results 49
   - 4.4. Remarks on Problems 54
   - 4.5. Conclusions 54

   References 57
5. The Language and Automation Bibliographic Project 58
6. Conclusions 67

   References 75

Appendix A: LINCS Document Characteristics Study 76

Appendix B: Examples of Types of Non-Textual Material Considered in Study of Language Sciences Primary Journals 82

Appendix C: Bibliography of Primary Works in a Sample of Computational-Linguistics Literature 89

LIST OF TABLES AND FIGURES

Tables

2. Comparison of Bibliographic Sources
   Table 1. Publications Scanned in Document-Characteristics Study 5
   Table 2. Comparison of Data Elements: Author Data 15
   Table 3. Comparison of Data Elements: Language Data 16
   Table 4. Comparison of Data Elements: Other Elements and Features 17
   Table 5. Comparison of Data Elements: Most Prominent Elements in Overall Graphic Representation 18
   Table 6. Variations in Journal-Citation Style 19
   Table 7. Comparison of Major Document Categories: Item in a Book 24
   Table 8. Comparison of Major Document Categories: Article in a Journal 25
   Table 9. Comparison of Major Document Categories: Book 26
   Table 10. Comparison of Major Document Categories: Book in a Series 27
   Table 11. Comparison of Major Document Categories: Book Review 28
   Table 12. Comparison of Major Document Categories: Conference Proceedings 29
   Table 13. Comparison of Major Document Categories: Thesis or Dissertation 30
   Table 14. Comparison of Major Document Categories: Foreign-Language Entry 31
   Table 15. Summary of Some Major Features of Bibliographic Practice in Secondary Sources Covering Language Science Literature 33
3. Some Typographic Aspects of the Primary Literature
   Table 1. Sample of Primary Journals 36
   Table 2. Percentages of Non-Textual Material in a Sample of Primary Journals 40
   Table 3. Percentage of Pages Containing Non-Textual Material 41
   Table 4. Results of the Special-Symbol Count 42
4. Requirements of the Character Set: A Sample of Literature in Computational Linguistics
   Table 1. Secondary Sources Scanned 46
   Table 2. References in Secondary Sources by Language 50
   Table 3. References in Periodicals by Language 51
   Table 4. References in Primary Works by Language 54
6. Conclusions
   Table 1. Analysis of References in Language Sciences Journals 71
   Table 2. Letters of the International Phonetic Alphabet 72

Figures

2. Comparison of Bibliographic Sources
Figure 1. Example of Coding Form for Type-Specific Data 8
Figure 2. Approaches to Graphic Accentuation of Bibliographic Elements 12
4. Requirements of the Character Set: A Sample of Literature in Computational Linguistics
Figure 1. Example of Work Sheet 47
Figure 2. Number of Unique Special Characters Required to Print a Given Percentage of Bibliographic References in the Primary and Secondary Literature in Computational Linguistics 56
5. The Language and Automation Bibliographic Project
Figure 1. Bibliographic Data Flow in LINCS Pilot Publication Program 61
Figure 2. Sample Completed Data Form 62
Figure 3. Entry Typed on MT/ST for Input to LINCS Data Base 63
Figure 4. Sample of Printout for Proofreading after Item has Passed through Input/Edit Stage 64
Figure 5. Mock-up of Printout from Publications Subfile; Tape Ready for Typesetter 65
Figure 6. The Typeset Entry in the Final Output 66
6. Conclusions
Figure 1. Distribution of Linguistic Literature by Type of Publication Outlet, Based on 1966 Linguistic Bibliography 70
Figure 2. Character Set of Roman-Alphabet and Transliterated Languages Used in Citations in the 1966 Linguistic Bibliography 73
Figure 3. Examples of Use of Special Characters in Linguistics Primary and Secondary Literature 74

Acknowledgments

The author would like to express her grateful appreciation to the following: Judith Krone, who conducted and described the study of typographic aspects of language sciences primary journals (section 3) and who provided valuable supervisory assistance in the coding and in the preparation of the tables for the bibliographic data (section 2); Stephen Christensen for coding and preparation of tables in the bibliographic characteristics study (section 2); Elizabeth Girrulat for all coding and Ludmila Okreglak for help with transliteration problems in the computational linguistics character set study (section 4) and in final proofing, Susan Paulus, Carl Wilson, and Charles Zisa for compilation of the list of the languages used in the Linguistic Bibliography; Alfred Pietrzyk, John Strawhorn, and the LINCS consultants for many helpful discussions and suggestions at various stages of this study, and Bernarda Erwin for constant help and encouragement and for patient editing of the report.

1. INTRODUCTION

A determination of the bibliographic and graphic characteristics of current research literature in the language sciences is essential to the design of an information system intended to offer primary and secondary publication services. Questions to be answered concern (1) the identification of documents per se, (2) their relative importance as information outlets, (3) the distinctions in the secondary literature, and (4) the constraints imposed by a network-based automated bibliographic system.

In connection with identification the questions are: What types of publications exist in the field and what are their distinguishing features? Are the latter physical features such as size, shape and "formal" appearance vs. prepublication form of a work? Or are the distinguishing features related to periodicity, function, or clientele?

Questions concerning distinctions in the secondary literature include the following: What publication types are typographically distinct in the secondary literature and how is this distinction achieved? What are the data elements used in their bibliographic representation? What is the minimum set of data elements necessary for the representation of the types of literature to be covered by a given bibliographic service?

In a consideration of constraints, the following are to be answered: Is the set of elements found in past and current services in the field satisfactory for search and retrieval requirements in an automated system? What limitations are or might be imposed by a network type of organization (i.e., the desirability of bibliographic information interchange); by document representation standards; and by typographic restraints in automated typesetting? On the basis of a study of past trends in bibliographic practices of various relevant services and on the basis of current requirements, can the set of elements and their most important configurations be described? Such information would serve as a guide in designing the information flow and in establishing the typographic requirements of a bibliographic information system. What are the graphic problems and requirements for computer recording, storing, composing, and printing of the diverse published material in the field? In primary linguistics journals, what are the prominent types of nontextual material that might pose special problems, and what is their approximate frequency of occurrence? In language sciences bibliographic publications, what are the requirements for special symbols resulting from the range of languages employed in citations? Cost, hardware constraints, and time factors involved in the development of special symbol sets also have to be borne in mind.

The study attempted to supply answers to some of these questions. Details of bibliographic practices of language sciences secondary services are outlined in section 2. The findings from this part of the study were utilized in the planning and design of an integrated approach to the collection, identification, computer manipulation, and display of bibliographic data in an experimental publication, Language and Automation. The Language and Automation project is described in section 5. Prominent typographic aspects and problems of primary core journals in linguistics and related fields are exemplified in section 3. Section 4 gives a first indication of the character set requirements in language sciences secondary publications. This was derived from examination of bibliographic references in a sample of bibliographies, journals, and primary works in a subfield of the language sciences, computational linguistics. As yet there are only tentative estimates of the overall need, in a linguistics publication system, for special alphabetic and other symbols. Comments on these, as well as on the relative importance of different document types in the language sciences, are found in section 6.

Among the problems which did not receive attention in this study were those relating to the identification and description, by physical characteristics and other relevant criteria, of the full range of formally- or semi-published information channels in the language sciences. Rather, the emphasis of this LINCS (Language Information Network and Clearinghouse System) was on the development of an automated bibliographic system.

2. COMPARISON OF BIBLIOGRAPHIC SOURCES

2.1. Objectives and Background

This review of the bibliographic practices of secondary publications relevant to the language sciences was undertaken as a preliminary to the design of LINCS bibliographic services. Of particular concern were the number, order, and graphic display of bibliographic data elements. Special attention was given to:

1. Distinct data elements employed in bibliographic records;
2. Distinct document types represented and their typographic image;
3. Variation among different bibliographies in representation of document categories and in general bibliographic features;
4. Languages used in bibliographic entries.

Primary attention was given to the graphic characteristics of language sciences documents in the secondary literature for several reasons. Authoritative and comprehensive secondary compilations are considered to reflect more or less realistically the range, form, and distribution of the literature in the field. Analysis of such sources is a less costly and time consuming approach to data collection than physical examination of the documents themselves [1]. The standardization of description already achieved in bibliographic records can be exploited, although perhaps at a loss in accuracy. The implications for network planning and bibliographic control of literature are also of importance. One of the basic premises of LINCS was that network-based sharing of bibliographic information was desirable and feasible. The planned application of the network concept by other developmental systems is often accompanied by a stated requirement for identification and detailed description of all functional elements which the bibliographic records to be handled in the system may contain (cf. Curran and Avram [2], Buckland [3]). To the extent that the bibliographic display of document categories is typographically or otherwise distinctive and can be described, processing, especially mechanized processing, and control of the information contained in the bibliographic record can be rendered more reliable. Mechanized control of bibliographic references for multipurpose use requires a more rigid documentation and detailed description of data elements and the relationships between them than is available in traditional bibliography operations. Such documentation and description is especially important if the secondary services to be rendered by the system are to cover the total range of the literature of a given field. Research projects in other scientific fields have therefore attempted to develop techniques for the description of bibliographic data elements that would allow reliable recognition and manipulation of the elements by computer [4, 5, 6]. Such explicit identification of data elements is the first step toward an automated publication system. In such a system, a typesetting program can translate functional identification codes into typographic specification for any desired output. Further, it was thought desirable that LINCS participate in the standardization efforts of such national and international organizations as the American National Standards Institute (ANSI), the National Federation of Science Abstracting and Indexing Services (NFSAIS), the International Organization for Standardization (ISO), and UNISIST, the World Science Information System. A final consideration was the importance of taking into account, wherever possible, current trends in language sciences publications concerning orthography, arrangement of bibliographic elements, typography, and other stylistic features.

2.2. Sources

The sources selected for study are listed in table 1 (see p. 5). They comprise 21 serial bibliographies and abstracting journals covering linguistics or language sciences topics, the journal Language, and three annual review publications. Language was included because it is the membership publication of the Linguistic Society of America and because it has a definitive style. Only the bibliographic references in the review journals and Language were examined. Although the most recent issues available for all sources were analyzed to insure that up-to-date style practices were reflected, the time range extended from 1964 (the latest issue for The Year's Work in English Studies) to 1969 (ERIC Research in Education).

2.3. Method

Since the study was limited to the bibliographic and stylistic practices of the source publications, it did not treat such matters as comprehensiveness and overlap in coverage, publication lags, geographic and subject coverage, etc., topics generally discussed in descriptions of the characteristics of the literature of a scientific discipline [7, 8].

A record of the maximally inclusive descriptive features of bibliographic citations for all document categories, that is, books, technical reports, proceedings, book reviews, articles, etc. would include the data elements themselves, the order in which the elements appeared, the type face and, if possible, point size distinctions the punctuation used to separate the elements, and any other stylistic or typographic features distinguishing one element from another (e.g. indentations, columnar arrangements, spacing, etc.). Since there is great variation in the representation of elements in different source bibliographies and within document categories, it proved difficult to develop a scheme or survey form of enough generality and yet sufficient detail to allow the reasonably faithful recording of all the peculiarities of any given bibliographic citation. Similar difficulties were reported in other studies [2]. The rather crude coding forms eventually arrived at provided for two sets of data. A source-specific form covered features characteristic of a given bibliography or abstract journal and independent of the types of entries included. On a type-specific form, data elements specific to the categories of documents represented by bibliographic entries were coded. Such types or categories were the physically distinctive classes of documents which might or might not be represented by differing sets and/or configurations of descriptive elements in diverse source bibliographies.

Table 1. Publications Scanned in Document-Characteristics Study

Publication	Abbreviation Used
Primary	
1. Language. Journal of the Linguistic Society of America	Language
Secondary	
2. Abstracts of English Studies	AbES
3. African Abstracts	AAb
4. L'Année Philologique: Bibliographie Critique et Analytique de l'antiquité Gréco-Latine	AP
5. Annual Bibliography of English Language and Literature	ELL
6. Bibliography of the Summer Institute of Linguistics	SIL
7. Bulletin Signalétique. 524: Sciences du Langage	BS
8. Computational Linguistics Bibliography (RAND)	CL
9. Computing Reviews	CRev
10. Current Index to Journals in Education	CIJE
11. dsh Abstracts	dshAb
12. ERIC Research in Education	RIE
13. The Finite String Bibliography	FS
14. Information Science Abstracts	ISAb
15. International Bibliography of the Social Sciences. Social and Cultural Anthropology	IBSC
16. LLBA: Language and Language Behavior Abstracts	LLBAb
17. Language-Teaching Abstracts	LTAb
18. A Language-Teaching Bibliography	LTB
19. Linguistic Abstracts (Linguistics)	LAb
20. Linguistic Bibliography	LB
21. MLA International Bibliography	MLA
22. Psychological Abstracts	PAb
Tertiary	
23. Annual Review of Information Science and Technology	RevIST
24. Biennial Review of Anthropology	RevA
25. The Year's Work in English Studies	RevES

For the sake of economy and speed in entering the descriptive data on the forms, codes were assigned to (1) the types of publications, (2) the elements found to represent these, and (3) the punctuation and type font associated with each element. Point sizes and other finer distinctions were eventually eliminated from consideration. Figure 1 is a typed version of a coded form. (In practice, these forms were completed by hand.) 'Source Code' identifies the bibliography being analyzed. 'Type of Publication' stands for one of the 32 classes or variations within classes identified.

Minor variations had to be coded separately since some bibliographies distinguished, either through a different arrangement or typographically, between entries for the category and the subcategory. In some bibliographies, for example, the author's name always appeared in the same position; in others the author preceded the title if an individual but followed if a corporate body. The types of publication were pragmatically defined as all those document types given some sort of unique treatment in the sources at hand, especially the Linguistic Bibliography. They included all the categories established by the ANSI preliminary Standard for Bibliographic References. The ANSI 'minimum of elements' was taken, where applicable, as the identifying criterion for a document class. The document categories and data elements used in the study are given in Appendix A.

A few additional remarks may illustrate the information obtained from the forms. 'Total Number' was intended to show how many data elements a given source bibliography employed to represent a given document type. [Naturally, this number was meaningful only if the example coded was typical of its kind and did not contain any incidental elements. This, however, was often not easy to determine. In the illustration in Figure 1, for example, the code 01 (for 'Entry number') under 'Name' is incidental to the type of publication coded, but entry numbers are characteristic of the particular source. In the analysis of such data only the minimal or typical elements for each type of publication were used and other elements ignored. In the absence of numbers of examples for a given type of publication it was often difficult to establish definitively the representativeness of the particular specimen found.]

Under 'Name' all elements of the bibliographic entry were coded; both elements of the citation proper and of descriptive matter (classification numbers, abstracts, etc.) were included. A total of 51 entry elements were used. These were derived in part from a preliminary examination of Linguistic Bibliography entries, partly from the ANSI Standard mentioned above, and partly from the list of data elements described by Curran and Avram [2] for the Committee on Machine Input Records of ANSI Z-39. The elements chosen did not, of course, include all of the possible components of descriptive and subject cataloging information found on Library of Congress cards, nor would they give all the information embodied in all the bibliographic entries. For

example, all bibliographic statements denoting publication stage or a special form of a work (e.g., manuscript, unpublished report, reprint, etc.) were coded as one element rather than separately, since the objective of the study was served by recording only those elements which might be graphically distinct or whose order within an entry might vary.

'Order in Entry' reflects the sequential order of data elements within an entry, elements being coded in the order of occurrence in a given bibliography.

The 'Type Font' in which elements were printed and the 'Punctuation' separating them were also coded.

Once the elements and document categories were defined, the actual coding did not present too many difficulties, especially in the case of source publications which in 'Instructions for use' or other front-matter specified the types of literature covered. For others, and in general for the bibliographies which covered the total range of the literature, considerable searching might be required. A time limit of 20-minutes was therefore set on searching for entries to establish the coverage of a certain document category in a given source publication or to establish the representativeness of the specimen entry found. The document categories or subcategories exemplified in entries in the source publications studied ranged from one (Abstracts of English Studies) to 22 (Linguistic Bibliography). The average was nine; one example for each was coded. An average of between four to six hours coding time was required for both the type-specific and the source-specific data for a given bibliography.

LINCS Document Characteristics Study

Type-Specific Data

Source: International Bibliography of the Social Sciences

Source Code	Type of Publication	Elements				Punctuation		Example 439
		Total Number	Name	Order in Entry	Type Font	Preceding	Succeeding	439 Lefebvre, H. Le langage et la societe. Paris, Gallimand, 66, 376 p. CR: M. Guyot, J. Soc. Americanistes 55(5), 1966: 273-275.
007	131		01	1	10	00	00	Comments/Explanations:
			04	2	31	02	03	Book reviews: the full entry is followed by CR (Compte Rendu), the name of the reviewer, and the review journal and collation. The name of reviewer is in caps and appears in the author index.
			09	3	20	03	03	
			26	4	10	03	04	
			27	5	10	04	04	
			29	6	10	04	04	
			31	7	10	04	03	
			42	8	31	03	04	
			41	9	20, 10	04	03	

Figure 1. Example of Coding Form for Type-Specific Data

2.4. Results and Comments

Tables 2-13 illustrate the type-specific and source-specific data coded. Features of the sources are described in tables 2-6 (pp. 15-19):

Table 2. Author data.
Table 3. Language data.
Table 4. Other elements and features.
Table 5. Most prominent elements in overall graphic representation.
Table 6. Variations in journal citation style.

32 document categories and subcategories were established for coding purposes, seven general classes are described in tables 7-13 (pp. 24-30):

Table 7. Item in a book.
Table 8. Article in a journal.
Table 9. Book.
Table 10. Book in a series.
Table 11. Book review.
Table 12. Conference proceedings.
Table 13. Thesis or dissertation.

Table 14. Foreign-language entry.

It should be noted that these findings do not necessarily reflect the publication policies of the source publications, they are rather a set of descriptive data reflecting the observed publication practices of bibliographic services covering linguistics. In general, the tables show the presence or absence of a given data element, its typography, and its order within the bibliographic citation. A hyphen indicates either that a given item is not relevant to a particular source bibliography as, for example, the element 'publisher' in a source covering only journal literature or that, if relevant, the source did not include this element.

A schematized summary of the most important bibliographic features and the findings regarding these features is presented in table 15.

2.4.1. Source Data

Table 2. Author Data. This table describes some aspects of the treatment of author names, for example, the order of presentation of last and given names, and the full or partial listing of multiple authors. There did not seem to be clear preference concerning the number of authors cited; inversion of names and spelling out were obviously favored.

Table 3. Language Data. In this table, 'Foreign-Language Title' refers to citations in any language other than the language of publication of the source bibliography. In most but not all cases they were the non-English entries.

The columns in this table were used as follows:

'Lang. Code' was marked if a statement identifying the language of the work was given for all foreign-language entries.

'All Titles in English' was marked for those sources which either covered only English-language publications or, while covering other-language material, did not give the title in the original language.

Under 'Foreign-Language Titles in Original Language', column 'All' was marked if all roman-script foreign-language entries found were given in the original. This did not imply that a given source bibliography actually covered the relevant literature in all such languages but only that for the literature covered the original language was used as a citation medium. 'Restricted Number' was marked if, in a few roman-alphabet language entries, a title translation rather than the original title was given. The four columns under 'Roman Alphabet' indicate which foreign materials were actually (or were thought to be) covered by a given source.

The non-roman alphabets found in the sources studied were Greek, Cyrillic, and Hebrew. The columns under 'Original Script' were used as follows: 'All' was marked if the source used these three scripts as citation media; 'Some' was marked if it used one or two; 'None' was marked if these scripts were transliterated and/or translated.

Some sources covered materials in a rather wide range of other non-roman alphabet languages (e.g. the Far-Eastern languages, Arabic, Bulgarian, Armenian, Persian), but did not publish entries in the original script. The 'Transliterated/Transcription' columns were marked 'All' if the languages used as media of citation were all transliterated; as 'Some', if some were transliterated, some only translated, the four sources marked as 'None' used the original scripts mentioned above but did not transliterate.

'Translation' columns were marked 'All' if title translations were consistently provided for foreign-language entries or for all except a few specified languages, as 'Some', if translations seem to be provided inconsistently and/or for only a few languages; as 'None', if foreign-language entries were given in the original language with no translation.

The language information contained in the table represents a general estimate of the range of foreign languages covered in the sources and the treatment of foreign-language citations. To determine precisely

through examination of each source the languages covered was not feasible in this study. It was attempted in the case of the Linguistic Bibliography in order to arrive at an estimate of character set requirements. The Linguistic Bibliography was found to employ over 30 languages, some 50 as citation media, the others in embedded phrases. It can be seen from the table, however, that in the great majority of sources studied foreign titles were given in the original language; in about one-third translations were given for all such titles. In another third, translations were given more or less consistently for most such titles but certain languages, usually specified, were excluded. Most roman-alphabet languages were represented as well as a considerable range of non-roman script languages in transliterated form. Of the non-roman alphabet languages, Greek, Cyrillic, and Hebrew were found in the original script.

Table 4. Other Elements and Features. On the assumption that publishing and pricing information were variables only in citations for books and other separately published material, the last five columns were blanked for all source publications whose coverage was restricted to journal articles. 'Journal Identification' refers to the fact that in a few sources place of origin of a journal might be included in all relevant entries, in addition to the full or abbreviated journal title.

Table 5. Most Prominent Elements. The two most prominent type face distinctions employed by a source were noted in order to illustrate the source's overall approach to bibliographic representation and the hierarchy of emphasis given to components of the citation. While the typographic variations did not necessarily indicate the relative importance ascribed to different elements, the way in which they made certain portions of the citation catch the reader's eye was as a rule characteristic of a given source. A graphic illustration is given in figure 2.

834.70

CARROLL, JOHN M.; DEBRUYN, JACOBUS G.

Univ. of Western Ontario, London (Canada), Dep. of Computer Science

On the importance of root-stem truncation in word-frequency analysis. ASIS (J Amer Soc Inform Sci), 1970, 21:5.368-369.

Research concerned with production of lists of keywords from text by statistical analysis. Objectives to determine if weighting coefficients would be adequate in fields different from that for which they were derived, establish whether extensive root-stem truncation procedure would improve performance, reduce computer time. Materials, experimental design, evaluation, conclusions. Five refs. – *From the author.*

Figure 2. Approaches to Graphic Accentuation of Bibliographic Elements (Reproduced with the permission of Language and Automation.)

In table 5 'Most Prominent' was defined as:

1. bold and caps (bo + c): any combination of type styles using both bold and caps;
2. bold (bo): any combination of type styles using bold;
3. caps (c): any combination of type styles not using bold;
4. italics (ital): any combination of type styles not using bold or caps;
5. roman (ro): only roman.

Roman was defined as any style or combination of styles in which neither bold, nor caps, nor italics was used. It is not limited strictly to the printer's roman types but may also include san serifs. Variants of types, such as different styles of roman type faces, and type size differences were not considered in this definition.

The treatment of some of the source publications in table 5 requires a word of explanation. In Abstracts of English Studies (AbES) individual entries consisted of entry number, author, title, pages, and abstract only. Individual entries were grouped under a collective journal title and collation (i.e. all items abstracted from a particular issue were given under the designation of that issue). In the journal designation the title was in bold and caps, volume, issue, and year were in bold. In the individual entries, the emphasis was on entry number and title. LLBA: Language and Language Behavior Abstracts was left blank in the table because its text, not typeset, appeared entirely in upper case. The Computational Linguistics bibliography (CL) was also not typeset but included in the table because typographic distinctions were indicated through underlining, etc. In Bulletin Signalétique (BS) the author was treated in two different ways. In non-review citations the author was given in caps and in inverted order; in book reviews the author was in roman and bold and in natural order. The Linguistic Bibliography (LB) and African Abstracts (AAb) employed typestyle variants for the author's surname and given name. In Psychological Abstracts (PAb) both author and title were in bold but the author's name was somewhat condensed with respect to the title and the rest of the entry.

Table 5 shows that 14 of the 25 sources studied used one or more boldface components. Entry number, author, title, descriptors, and order number appeared in boldface, with the title, and then the author, the most frequently boldfaced elements.

Table 6. Variations in Journal-Citation Style. Of all the bibliographic elements subject to variation in different sources, journal citations perhaps varied most. Table 6 illustrates the great diversification, from journal to journal, in bibliographic representation of these data; it also attempts to show each source's "basic" way of citing a serial published in volumes with regular issues and with continuous pagination over a volume. Most sources varied this style to reflect variations in the publication patterns of the journals cited; sources differed widely in the number of such distinctions made. Elements subject to variation were manner of publication (whether or not in regular volumes, with a specified and regular number of issues per volume); type of pagination (whether or not continuous over an entire volume); type of serial unit (self-contained and designated as number rather than volume); and similar items on which serials classifications might be based [9]. Although most sources used only one or two variations on a basic style, in others (e.g. the MLA International Bibliography) the modifications might be many.

Table 2. Comparison of Data Elements: Author Data

Source Publication	Personal Author					Corporate Author		
	Restriction on Number	All Inverted Order	All Natural Order	1st Inverted Others Natural	Given Name Initialized	Entry by Author	Entry by Title	Author Affiliation
AAb	yes	x			no			no
AbES				x	no			no
AP	yes	x			yes		x	no
BS	no	x			yes			no
CIJE	yes	x			no	x		no
CL	no		x		no			no
CRev	no	x			no		x	no
dshAb	no	x			yes	x		yes
ELL	yes			x	no	x		no
FS	no	x			no			no
IBSC	yes	x			yes	x		no
ISAb	no	x			no	x		yes
LAb			x		no			no
Language	no			x	no			no
LB	no	x			no		x	no
LLBAb	no	x			yes			yes
LTAb	no			x	no			no
LTB	yes			x	no	x		no
MLA	yes			x	no	x		no
PAb	yes	x			no	x		yes
RevA	yes			x	no	x		no
RevES	yes		x		no			no
RevIST	no	x			no	x		no
RIE	yes	x			no		x	yes
SIL	no			x	no			

Table 3. Comparison of Data Elements: Language Data

Source Publications	Lang. Code	All Titles in English: Only Eng. Pub.	All Titles in English: Foreign Lang. Pub.	Foreign-Language Titles in Original Language: All	Restricted Number	Roman Alphabet: Most Romance	Roman Alphabet: Most Germanic	Roman Alphabet: Most Slavic	Roman Alphabet: Some Slavic	Non-Roman Alphabets: Original Script: All	Original Script: Some	Original Script: None	Transliterated/Transcription: All	Transliterated/Transcription: Some	Transliterated/Transcription: None	Translation: All	Translation: Some	Translation: None
AAb				x		x	x		x			x	x			x		
AbES		x																
AP					x	x	x		x		x				x			x
BS					x	x					x				x		x	
CIJE		x																
CL				x								x	x			x		
CRev	x		x															
dshAb				x		x	x		x			x		x		x		
ELL				x			x		x		x			x				x
FS				x		x	x		x			x	x			x		
IBSC				x		x	x	x				x	x				x	
ISAb				x		x	x		x			x	x			x		
LAb				x		x	x	x				x	x					x
Language				x		x	x	x				x	x					x
LB				x		x	x	x			x			x			x	
LLBAb	x			x		x	x	x				x	x			x		
LTAb				x		x	x		x		x				x	x		
LTB				x		x					x				x		x	
MLA				x		x	x	x				x		x			x	
PAb				x		x	x					x	x			x		
RevA				x		x	x					x	x				x	
RevES		x																
RevIST		x																
RIE		x																
SIL				x								x	x				x	

Table 4. Comparison of Data Elements: Other Elements and Features

Source Publication	Capitalization of Title	Journal Title Abbreviation	Standard Abbreviation	Journal Identification	Place of Publication	Publisher	Year*	Pages	Price
AAb	no	yes	yes	no	-	-	-	-	-
AbES	yes	no	-	no	-	-	-	-	-
AP	no	yes	no	no	yes	yes	yes	yes	yes
BS	no	yes	yes	yes	yes	yes	yes	yes	yes
CIJE	yes	yes	yes	no	-	-	-	-	-
CL	yes	no	-	no	yes	yes	yes	no	no
CRev	no	yes	yes	no	yes	yes	yes	yes	yes
dshAb	no	yes	yes	no	yes	yes	yes	yes	no
ELL	yes	yes	no	no	yes	yes	no	yes	no
FS	yes	no	-	no	yes	yes	yes	no	no
IBSC	no	yes	yes	no	yes	yes	yes	yes	no
ISAb	no	no	-	no	yes	yes	yes	yes	yes
LAb	yes	no	-	no	yes	yes	yes	yes	yes
Language	no	no	-	no	yes	yes	yes	no	no
LB	no	yes	no	no	yes	yes	yes	yes	no
LLBAb	**	yes	yes	no	-	-	-	-	-
LTAb	no	no	-	yes	-	-	-	-	-
LTB	no	-	-	-	yes	yes	yes	yes	no
MLA	yes	yes	no	no	yes	yes	no	no	no
PAb	no	no	-	no	yes	yes	yes	yes	yes
RevA	no	yes	yes	no	yes	yes	yes	no	no
RevES	yes	yes	no	no	yes	yes	no	yes	yes
RevIST	no	no	-	no	yes	yes	yes	yes	no
RIE	**	no	-	no	yes	yes	yes	yes	yes
SIL	no	yes	no	no	yes	yes	yes	yes	no

*Listed as absent if the publication date is not explicitly given. In some source bibliographies the date for citations is not given, but is to be inferred from the year of publication of the source bibliography.

**Complete title in upper case.

Table 5. Comparison of Data Elements:
Most Prominent Elements in Overall Graphic Representation

Source Publication	Entry Number	Author	Title of Work	Title of Source	Year	Descriptors or Classification No.	Order Number
AAb	bo	bo+c	bo			bo	
AbES	bo	ro	bo				
AP		c,sc	ital				
BS		c	bo			bo, bo + ital	
CIJE			bo	ital			bo
CL			c	ital			
CRev		c,sc	bo				
dshAb	bo	bo+c	bo				
ELL	bo	c,sc					
FS		bo	*	ital			
IBSC		c.sc	*	ital			
ISAb			bo	ital			
LAb		ro	*	ital	ro		
Language		c,sc	ro	ro	ro		
LB		c,sc	*	ital			
LLBAb	-	-	-	-	-	-	-
LTAb	bo	bo		ital			
LTB	bo	bo	ital				
MLA		ro	*	ital			
PAb		bo	bo	ital			
RevA		ro	ro	ital	ro		
RevES		ro	ital	ital	ro		
RevIST		c	ro	ro	ro		
RIE			bo+c				bo
SIL		bo	ro	ro	ro		

*If the work is a book, the title is in italics; if an article, the title is in roman.

Table 6. Variations in Journal-Citation Style

Volume/issues, issues continuously paginated

AAb	Journal, **20**, 1964, 22-34.
AbES	JOURNAL, xix:4, Fall 1967.
AP	Journal xi 1964 343-365.
BS	Journal (1966), 21, 213-15.
CLJE	Journal v51, pp.69-76, Jan 69
CL	Journal, Vol. 1, No. 2 (1967), pp. 13-46.
CRev	Journal **8**, 8 (Aug. 1965), 501-508.
dshAb	Journal, 22(12), 1967, 119-125.
ELL	Journal, Spring, vii, 195-216.
FS	Journal, Vol. 25 (1964), pp. 274-282.
IBSC	Journal, 8(2), 1966 : 217-234.
ISAb	Journal **2**(4), 228-231 (1968 April).
LAb	Journal, 8, no. 3 (1966), 1-2.
Language	Journal 18.54-72.
LB	Journal XXX 1964, 156-169.
LLBAb	Journal, 1967, 55 (2), 179.
LTAb	Journal (The Hague), **91**, 3 (1967), 4-9.
LTB	-
MLA	Journal 7:28-31.
PAb	Journal, 1968, **13**(10), 793-798.
RevA	Journal 5: 1-20 (1964).
RevES	-
RevIST	Journal, 12 (July 1967) 344-345.
RIE	Journal; v1 n1 Nov 1967
SIL	Journal 3:205-26 (Jan 54)

2.4.2. Document-Category Data

Tables 7-14 exemplify for different document categories and for foreign-language entries the essential data elements, their type faces and order. The following abbreviations are used for type faces: c stands for caps, sc for small caps, b for boldface, ital for italics, and ro for any type face other than the above. Combinations of these such as, for example, c, b -- to stand for caps and bold -- are also used.

Table 7. Item in a Book. This document category generally represented contributions to collections like Festschriften, miscellanea, and proceedings, or chapters from books. The larger work in which the contribution appeared might or might not be the subject of a separate entry in the bibliography. The category is somewhat underrepresented in the table for several reasons. Only those sources were marked which gave complete or nearly complete bibliographic data for the larger work together with the citation of the contribution. Thus, sources like Psychological Abstracts and Computing Reviews, in which the reader was referred to a different entry for bibliographic data pertaining to the larger work, are left blank in the table. Source publications which, like the MLA International Bibliography, used a 'composite entry' approach in which titles of relevant items from a collection and the names of their authors were given in the abstract for, or as an annotation to, the citation for the collection itself, are not marked in the table. A variation of this document category -- where the collection was a special issue of a journal rather than a separately published work -- is not accounted for.

Table 8. Article in a Journal. Most bibliographies which covered both journal articles and articles from collections gave these categories the same graphic treatment. Within each citation the title of the item was distinguished from the title of its source by type style; in both citations the item titles were set in the same type and the source titles in a different one. Only two of the eleven sources used different type face combinations in these categories. For bibliographies which made typographic distinctions within the collation, the pattern most often found was boldface for the volume number and roman for the other parts of the collation.

Table 9. Book. The major distinction made in this category was in the typography of the title, source publications being almost evenly split between the use of boldface and italics. About a third of the sources included price in book citations.

Table 10. Book in a Series. Most of the source publications covering this category gave the series title and number immediately following the title of the work but in a different type style. If the series information did not appear immediately after the title, it was found at the end of the citation.

Table 11. Book Review. The table shows, as might be expected, that book reviews were covered mainly by annual bibliographies which attempted to cover comprehensively the literature of their respective fields. Source publications with a more restricted coverage -- eight in terms of document types referenced (e.g. Research in Education, Language Teaching Abstracts), or in terms of currency of publication (e.g. MLA International Bibliography, Language Teaching Bibliography) -- would not generally be expected to include book reviews. For comprehensive archival bibliographies like the Linguistic Bibliography, however, book review entries might constitute a substantial percentage of the total coverage. Book reviews might differ from normal book entries only in containing pertinent review data which was often introduced by 'Rev. by' or a similar identifying phrase, or they might be typographically distinct as in Bulletin Signalétique or the Linguistic Bibliography. In the former, the author was given in boldface as opposed to caps for all other entries; in the latter, the review entry number was italicized while the other entry numbers were in times roman.

The reviewer's name did not generally appear in the author indices. In the International Bibliography of the Social Sciences and Language and Language Behavior Abstracts, however, reviewers were listed along with authors.

Table 12. Conference Proceedings. This category was treated in several ways. Entries were grouped in a special section on conferences. Or, where a brief title incorporated the subject or theme of a conference, it was treated as the title of the entry, with all conference particulars like date, sponsorship, etc. following as in a series statement. In most cases, the type styles used for these elements corresponded to those for a book in a series. In other sources, the title was always given as the first element in the citation.

Table 13. Thesis or Dissertation. There were two major variations in the representation of this document category. The thesis or dissertation was treated as a separately published work. The typography was generally that used for a book in a series, with identification of type of degree and degree-granting institution taking the place of a series statement, and abstracting or reviewing journal data appended as in a book review. 12 of the 14 sources listed use this approach. The thesis or dissertation was sometimes treated typographically as a journal article. Identification as a dissertation might be by implication only (the source journal being Dissertation Abstracts) or the degree-granting institution might be given after the source journal citation.

Table 14. Foreign-Language Entry. This table exemplifies foreign-language citations for separately published works such as books and for journal articles. Generally, an example for a book is given. However, for those source publications which cover only periodical literature, 'Title of Journal' takes the place of 'Publishing Data' and full 'Collation' is implied by 'Year and Pages'. More than one typeface in 'Publishing Data' and 'Year and Pages' indicates typographic differentiation within the journal collation data (see table 6).

While most of the source publications analyzed covered foreign-language material and routinely provided translations, in some cases the examples described in table 14 were the exception rather than the rule, either because the subject orientation was restricted to English studies or because translations were found for only a very restricted number of foreign languages. Foreign-language entries were infrequent in Research in Education and Annual Review of Information Science and Technology, references in the journal Language were almost always without translation (the example described was a transliterated Japanese entry); and the Language-Teaching Bibliography translated only Russian titles (printed in Cyrillic). In addition, there was some inconsistency and wide latitude in the number of languages for which a source bibliography

provided translations. The Linguistic Bibliography supplied translations for titles in all languages other than French, English, German, Spanish, Portuguese, and Italian. The MLA International Bibliography listed titles in many languages but as a rule did not translate those in English, French, German, Spanish, Italian, Portuguese, Romanian, Scandinavian, Dutch, Celtic; ten additional East European languages were not translated if they appeared within their respective language sections, but if a title in Russian, for example, appeared in a general section of the Bibliography a translation might be provided. The Annual Bibliography of English Language and Literature generally gave a transliterated and original title for all East European languages, but there were occasional entries with only a translated title and a statement indicating the language of the original.

Table 7. Comparison of Major Document Categories: Item in a Book

Source Publication	Author		Title of Item		Editor of Book		Title of Book		Place of Publication		Publisher		Year of Publication		Pages	
	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order
AAb	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
AbES	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
AP	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
BS	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
CIJE	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
CL	ro	2	c	1	ro	3	ital	4	ro	6	ro	5	ro	7	ro	8
CRev	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
dshAb	c,b	1	b	2	ro	3	ro	4	ro	5	ro	6	ro	7	ro	8
ELL	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
FS	b	1	ro	2	–	–	ital	3	ro	5	ro	4	ro	6	ro	7
IBSC	c,sc	1	ro	2	c,sc	3	ital	4	ro	5	ro	6	ro	7	ro	8
ISAb	ro	1	b	2	ro	3	ital	4	ro	7	ro	6	ro	5	ro	8
LAb	ro	1	ro	2	–	–	ital	3	ro	4	ro	5	ro	6	ro	7
Language	c,sc	1	ro	3	ro	5	ro	4	ro	7	ro	8	ro	2	ro	6
LB	c,sc	1	ro	2	–	–	ital	3	ro	4	ro	5	ro	6	ro	7
LLBAb	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
LTAb	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
LTB	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
MLA	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
PAb	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
RevA	ro	1	ro	2	ro	3	ro	4	ro	5	ro	6	ro	7	–	–
RevES	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
RevIST	c	1	ro	2	ro	3	ro	4	ro	6	ro	5	ro	7	ro	8
RIE	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
SIL	b	1	ro	2	ro	3	ro	4	ro	5	ro	6	ro	7	ro	8

Table 8. Comparison of Major Document Categories: Article in a Journal

Source Publication	Author		Title		Journal				Journal Identification		Vol./Issue		Pages		Year	
					Full Title		Abbreviation									
	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order
AAb	c,b	1	b	2	-	-	ro	3	-	-	ro,b ro	4	ro	6	ro	5
AbES	ro	4	b	5	c,b	1	-	-	-	-	ro,b	2	ro	6	ro,b	3
AP	c,sc	1	ital	2	-	-	ro	3	-	-	ro	4	ro	6	ro	5
BS	c	1	b	2	-	-	ital	3	ital	4	ital ro	6	ro	7	ro	5
CIJE	ro	2	b	1	-	-	ital	3	-	-	ro	4	ro	5	ro	6
CL	ro	2	c	1	ital	3	-	-	-	-	ro	4	ro	6	ro	5
CRev	c,sc	1	b	2	-	-	ital	3	-	-	ro,b ro	4	ro	6	ro	5
dshAb	c,b	1	b	2	-	-	ital	3	-	-	ro	4	ro	6	ro	5
ELL	c,sc	1	ro	2	-	-	ro	3	-	-	ro	5	ro	6	-	-
FS	b	1	ro	2	ital	3	-	-	-	-	ro	4	ro	6	ro	5
IBSC	c,sc	1	ro	2	-	-	ital	3	-	-	ro	4	ro	6	ro	5
ISAb	ro	1	b	2	ital	3	-	-	-	-	ro	4	ro	5	ro	6
LAb	ro	1	ro	2	ital	3	-	-	-	-	ro	4	ro	6	ro	5
Language	c,sc	1	ro	3	ro	4	-	-	-	-	ro	5	ro	6	ro	2
LB	c,sc	1	ro	2	-	-	ital	3	-	-	ro	4	ro	7	ro	6
LLBAb	-	1	-	2	-	-	-	3	-	-	-	5	-	6	-	4
LTAb	b	1	ro	2	ital	3	-	-	ro	4	ro,b ro	5	ro	7	ro	6
LTB	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MLA	ro	1	ro	2	-	-	ital	3	-	-	ro	4	ro	5	-	-
PAb	b	1	b	2	ital	3	-	-	-	-	ro,b ro	5	ro	6	ro	4
RevA	ro	1	ro	2	-	-	ital	3	-	-	ro	4	ro	5	ro	6
RevES	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RevIST	c	1	ro	2	ro	3	-	-	-	-	ro	4	ro	6	ro	5
RIE	ital	1	c,b	2	ro	3	-	-	-	-	ro	4	ro	5	ro	6
SIL	b	1	ro	2	-	-	ro	3	-	-	ro	4	ro	5	ro	6

Table 9. Comparison of Major Document Categories: Book

Source Publication	Author		Title		Edition		Place of Publication		Publisher		Year of Publication		Pages		Price	
	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order
AAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AbES	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AP	c,sc	1	ital	2	ro	3	ro	4	ro	5	ro	6	ro	7	ro	8
BS	c	1	b	2	ro	3	ro	4	ro	5	ro	6	ro	7	ro	8
CIJE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CL	ro	2	c	1	-	-	ro	4	ro	3	ro	5	-	-	-	-
CRev	c,sc	1	b	2	ro	3	ro	5	ro	4	ro	6	ro	7	ro	8
dshAb	c,b	1	b	2	-	-	ro	3	ro	4	ro	5	ro	6	-	-
ELL	c,sc	1	ro	2	ro	3	ro	4	ro	5	-	-	ro	6	-	-
FS	b	1	ital	2	-	-	ro	4	ro	3	ro	5	-	-	-	-
IBSC	c,sc	1	ital	2	ro	3	ro	4	ro	5	ro	6	ro	7	-	-
ISAb	ro	1	b	2	-	-	ro	5	ro	4	ro	3	ro	6	-	-
LAb	ro	1	ital	2	ro	3	ro	4	ro	5	ro	6	ro	7	ro	8
Language	c,sc	1	ro	3	ro	4	ro	5	ro	6	ro	2	-	-	-	-
LB	c,sc	1	ital	2	ro	3	ro	4	ro	5	ro	6	ro	7	-	-
LLBAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LTAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LTB	b	1	ital	2	ro	5	ro	3	ro	4	ro	6	ro	7	-	-
MLA	ro	1	ital	2	ro	5	ro	3	ro	4	-	-	-	-	-	-
PAb	b	1	b	2	-	-	ro	3	ro	4	ro	5	ro	6	ro	7
RevA	ro	1	ro	2	-	-	ro	3	ro	4	ro	5	-	-	-	-
RevES	ro	2	ital	1	ro	3	-	-	ro	4	-	-	ro	5	ro	6
RevIST	c	1	ro	2	ro	3	ro	5	ro	4	ro	6	ro	7	-	-
RIE	ital	1	c,b	2	-	-	ro	6	ro	5	ro	3	ro	4	ro	7
SIL	b	1	ro	2	ro	3	ro	4	ro	5	ro	6	ro	7	-	-

Table 10. Comparison of Major Document Categories: Book in a Series

Source Publication	Author		Title of Work		Title & No. of Series		Editor of Series		Place of Publication		Publisher		Year		Pages		Price	
	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order
AAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AbES	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AP	c,sc	1	ital	2	ro	3	-	-	ro	4	ro	5	ro	6	ro	7	ro	8
BS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CIJE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CL	ro	2	c	1	ro	3	-	-	ro	5	ro	4	ro	6	-	-	-	-
CRev	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
dshAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ELL	c,sc	1	ro	2	ro	6	-	-	ro	3	ro	4	-	-	ro	5	-	-
FS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IBSC	c,sc	1	ital	2	ro	7	-	-	ro	3	ro	4	ro	5	ro	6	-	-
ISAb	ro	1	b	2	ital	3	-	-	ro	6	ro	5	ro	4	ro	7	ro	8
LAb	ro	1	ital	2	ital	3	-	-	ro	4	ro	5	ro	6	ro	7	ro	8
Language	c,sc	1	ro	3	ro	4	-	-	ro	5	ro	6	ro	2	-	-	-	-
LB	c,sc	1	ital	2	ro	3	-	-	ro	4	ro	5	ro	6	ro	7	-	-
LLBAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LTAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LTB	b	1	ital	2	ro	3	-	-	ro	4	ro	5	ro	6	ro	7	-	-
MLA	ro	1	ital	2	ro	3	-	-	ro	4	ro	5	ro	6	-	-	-	-
PAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RevA	ro	1	ro	2	ro	3	-	-	ro	4	ro	5	ro	6	-	-	-	-
RevES	ro	2	ital	1	ro	3	-	-	ro	4	ro	5	-	-	ro	6	ro	7
RevIST	c	1	ro	2	ro	7	-	-	ro	4	ro	3	ro	5	ro	6	-	-
RIE	ital	1	c,b	2	c,b	3	-	-	ro	7	ro	6	ro	4	ro	5	ro	8
SIL	b	1	ro	2	ro	3	-	-	ro	4	ro	5	ro	6	ro	7	-	-

Table 11. Comparison of Major Document Categories: Book Review

Source Publication	Author		Title		Place of Publication		Publisher		Year		Pages		Price		Review		Reviewer	
	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order
AAb	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
AbES	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
AP	c,sc	1	ital	2	ro	3	ro	4	ro	5	ro	6	ro	7	ro	8	ro	9
BS	b	1	b	2	ro	6	ro	5	ro	7	ro	8	–	–	ital	3	c	4
CIJE	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
CL	ro	2	c	1	ro	4	ro	3	ro	5	–	–	–	–	ital	7	ro	6
CRev	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
dshAb	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
ELL	c,sc	1	ro	2	ro	3	ro	4	–	–	ro	5	–	–	ro	7	ro	6
FS	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
IBSC	c,sc	1	ital	2	ro	3	ro	4	ro	5	ro	6	–	–	ital	8	c,sc	7
ISAb	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
LAb	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Language	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
LB	c,sc	1	ital	2	ro	3	ro	4	ro	5	ro	6	–	–	ital	7	ro	8
LLBAb	–	3	–	2	–	4	–	5	–	6	–	–	–	–	–	7	–	1
LTAb	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
LTB	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
MLA	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
PAb	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
RevA	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
RevES	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
RevIST	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
RIE	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
SIL	b	1	ro	2	ro	3	ro	4	ro	5	ro	6	–	–	ro	8	ro	7

Table 12. Comparison of Major Document Categories: Conference Proceedings

Source Publication	Title (Theme) of Conference		Conference Particulars		Editor (or) Author		Place of Publication		Publisher		Year		Pages		Price	
	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order
AAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AbES	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AP	ital	2	ro	3	c,sc	1	ro	4	ro	5	ro	6	ro	7	-	-
BS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CIJE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CL	c	1	ro	3	ro	2	ro	5	ro	4	ro	6	-	-	-	-
CRev	b	2	ro	3	c,sc	1	ro	5	ro	4	ro	6	ro	7	-	-
dshAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ELL	ro	1	-	-	-	-	ro	2	ro	3	-	-	ro	4	-	-
FS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IBSC	ital	2	-	-	c,sc	1	ro	3	ro	4	ro	5	ro	6	-	-
ISAb	b	2	b	3	ro	1	ro	6	ro	5	ro	4	ro	7	-	-
LAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Language	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LB	ital	1	ro	2	ro, c,sc	3	ro	4	ro	5	ro	6	ro	7	-	-
LLBAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LTAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LTB	ital	2	ro	3	b	1	ro	4	ro	5	ro	6	ro	7	-	-
MLA	ital	2	ital	3	ro	1	ro	4	ro	5	ro	6	-	-	-	-
PAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RevA	ro	2	-	-	ro	1	ro	3	ro	4	ro	5	-	-	-	-
RevES	ital	1	-	-	ro	2	ro	3	ro	4	-	-	ro	5	ro	6
RevIST	ro	2	ro	3	c	1	ro	5	ro	4	ro	6	ro	7	-	-
RIE	c,b	1	ro	4	-	-	ro	6	ro	5	ro	2	ro	3	-	-
SIL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 13. Comparison of Major Document Categories: Thesis or Dissertation

Source Publication	Author		Title		Institution		Degree		Year		Pages		Abstracting Journal	
	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order
AAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AbES	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AP	c,sc	1	ital	2	ro	4	ro	3	ro	5	ro	6	ro	7
BS	c	1	b	2	-	-	ro	6	ro	4	ro	5	ital	3
CIJE	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRev	c,sc	1	b	2	ro	4	ro	3	ro	5	-	-	-	-
dshAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ELL	c,sc	1	ro	2	ro	4	ro	3	*-	-	-	-	ro	5
FS	b	1	ital	2	ro	4	ro	3	ro	5	-	-	-	-
IBSC	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ISAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LAb	ro	1	ital	2	ro	4	ro	3	ro	5	ro	6	-	-
Language	c,sc	1	ro	3	ro	5	ro	4	ro	2	-	-	-	-
LB	c,sc	1	ital	2	ro	4	ro	3	ro	5	ro	6	ital	7
LLBAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LTAb	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LTB	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MLA	ro	1	ro	2	ro	4	-	-	*-	-	-	-	ital	3
PAb	b	1	b	3	ro	2	-	-	-	-	-	-	ital	4
RevA	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RevES	ro	2	ital	1	ro	4	ro	3	*-	-	ro	5	-	-
RevIST	c	1	ro	2	ro	4	ro	3	ro	5	-	-	-	-
RIE	ital	1	c,b	2	ro	6	ro	5	ro	3	ro	4	-	-
SIL	b	1	ro	2	ro	4	ro	3	ro	5	ro	6	-	-

*year is implied

Table 14. Comparison of Major Document Categories: Foreign-Language Entry

Source Publication	Author		Title of Work		Title Translation		Publishing Data		Year and Pages	
	Type	Order	Type	Order	Type	Order	Type	Order	Type	Order
AAb [periodical]	c,b	1	b	2	ro	3	ro	4	ro,b	5
AbES	-	-	-	-	-	-	-	-	-	-
AP [book]	c,sc	1	ital	2	-	-	ro	3	ro	4
BS [book]	c	1	b	2	ro	3	ro	4	ro	5
CIJE	-	-	-	-	-	-	-	-	-	-
CL [book]	ro	2	ital	3	c	1	ro	4	ro	5
CRev [book]	c,sc	1	-	-	b	2	ro	3	ro	4
dshAb [periodical]	c,b	1	b	2	ro	3	ital	4	ro	5
ELL [book]	c,sc	1	ro	2	-	-	ro	3	ro	4
FS [book]	b	1	ital	2	ital	3	ro	4	-	-
IBSC [book]	c,sc	1	ital	2	ro	3	ro	4	ro	5
ISAb [book]	ro	1	b	2	ro	3	ro	5	ro	4
LAb [book]	ro	1	ital	2	-	-	ro	3	ro	4
Language [book]	c,sc	1	ro	3	ro	4	ro	5	ro	2
LB [book]	c,sc	1	ital	2	ro	6	ro	3	ro	5
LLBAb [periodical]	-	1	-	2	-	3	-	4	-	5
LTAb [periodical]	b	1	ro	2	ro	3	ital	4	ro	5
LTB [book]	b	1	ital	2	ro	3	ro	4	ro	5
MLA [book]	ro	1	ital	2	ro	4	ro	3	-	-
PAb [book]	b	1	b	2	ro	3	ro	4	ro	5
RevA [book]	ro	1	ro	2	-	-	ro	3	ro	4
RevES [book]	ro	2	ital	1	-	-	ro	3	ro	4
RevIST [book]	c	1	ro	2	ro	3	ro	4	ro	5
RIE [book]	ital	1	c,b	2	c,b	3	ro	4	ro	5
SIL [book]	b	1	ro	2	ro	3	ro	4	ro	5

2.5. Summary and Conclusion

The results of this comparative study of bibliographic practices in language sciences secondary publications are schematized in table 15. This table summarizes some of the more important features of bibliographic practice in secondary services and gives approximate percentages of occurence in the sample. Sources constituting the relatively small, very heterogeneous sample differed widely in their document coverage, use of foreign languages and 'centrality' to the language sciences. The features selected for consideration were not always prominent in the sample and possibly reflected the study's biases concerning their 'importance' in practical bibliographic considerations. It may be said that the bibliographic services covering the language sciences, in the aggregate, show most of the data elements which current documentation standards and groups concerned with bibliographic information interchange define as essential for description of printed records. Individual bibliographies differ widely concerning not only the number of document categories covered but also the completeness of bibliographic entries. An 'ideal' model for bibliographic description which meets standards requirements can therefore be derived only from the total set of features and elements. Important bibliographic standards to be considered in the determination of a final set of elements include those of ISO Subcommittee TC46, ANSI Z-39 Subcommittee 4 on Bibliographic References, and the set of data elements of the Library of Congress MARC Project. The recommendations and standardization efforts of UNISIST and the Bibliographic Citation Committee of the National Federation of Science Abstracting and Indexing Services (NFSAIS) are also relevant. The composite of discrete items obtained in the study contains very nearly all the data elements which are 'essential' or 'recommended' by these groups. The results of the study were put to use in the design of the data structure and data forms for the LINCS reference publication, Language and Automation. The bibliographic automation pilot project under which this publication is produced is described in section 5.

In addition to standards determining the adequacy of the set of descriptive data elements, other standards relating to treatment of individual elements or document categories should be considered. This study has not attempted however to establish the degree to which language sciences sources conform to Z-39 standards on title word abbreviations, country and language name abbreviations, COSATI standards on technical reports, etc. It is possible that there is less conformity in this regard than in the case of data elements. Many of these standards, however, become increasingly important in the planning for comprehensive information systems with automated search and publication services and data interchange with other systems. In this instance, past practices of printed information services, especially in a field with a strong humanistic orientation, may be less important than an awareness of the relevant standards and their applicability to the functioning of an integrated system.

Table 15. Summary of Some Major Features of Bibliographic Practice in Secondary Sources Covering Language Sciences Literature

Entry elements emphasized (Tables 4, 5, 7-14)	Title distinguished typographically 36%	Author distinguished typographically 48%	Author precedes title 88%	Title distinguished orthographically-capitalization of nouns 35%
Author treatment (Table 2)	Restriction on number listed 46%	Inverted order for all authors 64%	Given name spelled out 75%	Affiliation given 25%
Source journal treatment (Tables 4, 6)	Title abbreviated 67%	Standard abbreviation 59%	Year first element in collation 13%	Volume, number, part, or page designation explicit 21%
Foreign language use (Tables 3, 14)	Foreign titles in original language 68%	Non-roman alphabets transliterated 68%	Translation provided 64%	Foreign title precedes translation 80%
Other entry elements (Tables 4, 11)	Place of publication given 100%	Publisher given 85%	Year of publication given explicitly 85%	Book review data with item reviewed 100%

REFERENCES

1. Creager, William A.; et al. A serials data program for science and technology. Results of a feasibility study. Reading, Mass.: Information Dynamics, 1965.

2. Curran, Ann T.; Avram, Henriette D. The identification of data elements in bibliographic records. Washington, D.C.: Library of Congress, 1967.

3. Buckland, Lawrence F. Problems of recording text information in machine form for use in a scientific information communication network. Maynard, Mass.: Inforonics, 1966.

4. Thompson, Margaret. Automatic reference analysis. Scientific report ISR-3. Cambridge, Mass.: Harvard University, 1963.

5. Sawin, Lewis Clard, Roger, eds. Integrated bibliography pilot study. Progress report. Boulder, Colo.: Integrated Bibliography Pilot Study, 1965.

6. Blum, Arthur R. Data element standardization as an aid in the transfer of physics information. Proceedings of the American Society for Information Science 5 (New York: Greenwood, 1968), 117-119.

7. Creager, William A. A study of the current characteristics of scientific literature in the field of mental health. Report to the National Institute of Mental Health. Washington, D.C.: Herner, 1963.

8. Hawkes, H. E. The literature of geology. A study of scientific publications in the geological sciences. NSF Contract C-310. Washington, D.C.: 1966.

9. Bishop, David; Milner, Arnold L.; Roper, Fred W. Publication patterns of scientific serials. American Documentation, 1965, 16:2.113-121.

3. SOME TYPOGRAPHIC ASPECTS OF THE PRIMARY LITERATURE

3.1. Purpose

General typographic aspects of primary journals in the language sciences were examined to determine the approximate percentage of non-textual material in a sample of primary journals and to estimate the average number of special symbols used per page in each journal of the sample.

3.2. Method

3.2.1. Sample Selection

A sample of 15 journals was selected. Nine journals were chosen from a list of the most important core journals in general linguistics as determined by a questionnaire previously distributed among linguists at the Center for Applied Linguistics [1]. This group of journals was supplemented by an additional six journals taken from the bibliographical handbook Information Sources in Linguistics [2] in an attempt to give representation to applied linguistics and to various disciplines related to linguistics. Journals not published in conventional printed form or published in non-roman script were avoided. The complete 1967 volume for each journal was used in most cases; where several volumes were published for 1967, only one or two were included. If the volume contained a special issue (e.g. an issue devoted entirely to book reviews), that issue was disregarded. All pages except title pages, tables of contents, and volume indices were included in the total number of pages per journal and examined for non-textual material. The titles of the 15 journals in the sample are given in table 1, together with their abbreviations as listed in the Linguistic Bibliography [3].

Table 1. Sample of Primary Journals

Abbreviation	Title
BSL	Bulletin de la Société de Linguistique de Paris
CJL	Canadian Journal of Linguistics
FL	Foundations of Language
IJAL	International Journal of American Linguistics
JL	Journal of Linguistics
---	Language
L&S	Language and Speech
---	Lingua
--	Linguistics
CAnthr	Current Anthropology
CLTA	Cahiers de Linguistique Théorique et Appliquée
JVLVB	Journal of Verbal Learning and Verbal Behavior
MLJ	Modern Language Journal
MT	Mechanical Translation
--	Phonetica

3.2.2. Classification of Non-Textual Material

Nine different types of non-textual material were selected for consideration: tables, special tables, graphs, tree structures, mathematical expressions, maps, photographs, references, and advertising.

Since there were some items which did not fit specifically into any one category and since a large miscellaneous category was considered undesirable, it was necessary to broaden several categories as indicated below. Items referred to as set off from the text were found separated from the body of the printed page by brackets or braces, by spacing, by change in type font or size, or by some combination of these.

1. 'Tables' include items specifically labelled tables or charts as well as items set off from the text that are boxed in and/or arranged in columnar form and vowel or consonant patterns set off from the text.

2. 'Special Tables' include tables run lengthwise across the page, fold-out tables, and two-dimensional tables.

3. 'Graphs' represent line graphs, bar graphs, flow charts, and all other linear diagrams except tree structures.

4. 'Mathematical Expressions' cover not only equations and standard mathematical expressions but also statements set off from the text employing common mathematical symbols (e.g. $\sqrt{}$, subscripts, superscripts, etc.)

5. 'References' consist of footnotes as well as lists of references.

Examples of the types of items included in the categories 'Tables', 'Special Tables', 'Graphs', 'Tree Structures', and 'Mathematical Expressions' are provided in Appendix B.

3.2.3. Determining the Size of an Item

The size of any non-textual item was considered to be the percentage of the page which it occupied, including the space immediately above and below as well as captions or figure statements. Approximate percentage was generally estimated to the nearest 10% on a scale from 10 to 100%. The number of columns and the average number of lines on a full page of print for any given journal were taken into account in determining percentage. Frequently several items in the same category on a page were considered as a single item in the calculation of the percentage for

that page; therefore, a count of the specific number of tables, graphs, mathematical expressions, etc. was not made. Non-textual elements such as graphs and tables embedded in footnotes occurred relatively infrequently and were counted only once, in the 'Reference' category as part of the footnote space.

When for every page in a journal the percentage of non-textual material had been classified as to category, the number of pages each category occupied was determined, and, using the total number of pages in the journal, the percentage of journal space was calculated for each category (see table 2, p. 40). These figures, in effect, represented the space each category would occupy if its contents were grouped together without intervening text. The items in each category were of course scattered throughout the text of the journal. Since cost would normally be figured per page, table 3 (see p. 41) gives for each journal the percentage of the total number of pages that contained any non-textual material.

3.2.4. Symbol Count

To get an idea of how frequently certain types of symbols were used, the number of special symbols on every tenth page of each journal was counted. Title pages, tables of contents, indices, and pages of advertising were disregarded. If the tenth page was not a full page of print, the next full page was taken and counting continued from there. The term 'special symbol' includes letters with diacritics, letters from non-roman alphabets, phonetic symbols that differ from ordinary letters of the roman alphabet, subscripts, superscripts, and other mathematical symbols (excluding +, -, ?, ', =, /, $, ¢, #, %, *). In general, the idea was to include those symbols which were relatively uncommon and might require an extended character set. Special symbols found in mathematical expressions, tree structures, tables, footnotes, reference lists, and figure statements for graphs were covered in the count. Selections of text reflecting the variety of special symbols used in the journals are given in Appendix B.

3.3. Results

Table 2 gives the percentages rounded to the nearest .1% of non-textual material in various categories for each journal. The total percentage of non-textual material for the sample ranges from 13.7% to 37.9%, with an average of 22.9%.

Table 3 gives the abbreviated title for each journal, the total number of pages in the 1967 volume, and the number and percentage of pages which contain any quantity of non-textual material in any of the nine categories defined. 44% to 79% of the total number of pages in each journal contained some non-textual material with 68% being the figure that occured most frequently. Since even pages with one mathematical expression or a single footnote were counted, these figures are intended in each case to represent a maximum estimate.

The results of the rough count of special symbols are given in table 4 (see p. 42). The abbreviated titles are listed in descending order beginning with the journal having the greatest average number of symbols per page. The 'Language(s)' column gives, for each journal, the languages in which book reviews and articles in the 1967 volume appear. In several instances where, for any journal, the difference between the highest number of symbols per page and the next highest number was very great, the highest figure was dropped from the range in order to reduce the distortion in the average that would result from a positively skewed distribution. A page printed in Cyrillic in the Journal of Linguistics had over 1,000 symbols; the page with the next highest number of symbols had 119, and the other pages fell substantially below that figure. Including the page of Cyrillic the JL average number of symbols per page is 54; without it, the average is only 8.

The average number of symbols per page in the sample ranged from three to 95. Differences between journals in this regard reflected in part the differences in the languages which appeared and in the types of articles published. A journal which published articles only in French (BSL) would have a high average symbol count because of the number of letters with accent marks. A journal largely devoted to phonological aspects of languages (IJAL) would have a high average symbol count due to the frequent use of phonetic symbols.

It must be kept in mind that the results of this substudy represent only rough estimates. With regard to non-textual material, percentages are only approximate since graphs which were actually 12% or 14% of a page were counted as 10%; tables representing 15% or 17% of a page, as 20%, etc. Furthermore, when 10% of a page of print consisted of 3-4 lines, the difference between a 10% graphic and a 20% graph was slight and the percentage difference rather misleading.

The procedure of counting the special symbols on every tenth page resulted in the examination of very few pages in those journals with a small number of pages. Also, symbols, instead of being evenly distributed, appeared to be concentrated in certain articles (e.g. articles written in foreign languages, articles applying mathematical concepts to languages, etc.) which might have been missed completely if only every tenth page was considered.

Table 2. Percentages of Non-Textual Material in a Sample of Primary Journals

	BSL	CJL	FL	IJAL	JL	Language	L&S	Lingua	Linguistics	CAnthr	CLTA	JVLVB	MLJ	MT	Phonetica
Tables	3.6	4.2	3.5	6.2	6.0	5.1	14.8	4.1	6.1	10.0	18.7	6.7	3.6	14.7	7.3
Special Tables	3.6	3.1	--	.3	1.5	.4	8.2	--	--	--	1.1	.7	.7	--	.8
Graphs	.1	.8	.2	.2	1.5	1.2	8.2	1.0	.6	1.4	2.7	3.7	.2	9.3	5.2
Tree Structures	--	.2	2.1	.3	2.0	1.8	.1	.4	.9	.1	4.1	--	--	--	.4
Mathematical Expressions	.6	2.4	2.7	.7	2.8	2.4	--	3.7	.7	.1	6.0	.2	--	2.3	.6
Maps	.5	5.6	--	--	--	.3	--	--	--	.4	--	--	--	--	--
Photographs	--	--	--	--	--	--	.4	--	1.1	3.0	--	.1	--	--	1.6
References	8.2	7.9	3.9	6.0	3.4	6.6	6.2	6.2	8.0	6.7	3.5	5.6	1.0	3.3	5.4
Advertising	--	.9	2.7	--	3.3	--	--	--	--	--	--	2.3	31.0	--	--
Total %	16.6	24.9	15.1	13.7	20.5	17.8	37.9	15.4	17.4	21.7	36.1	19.3	36.5	29.6	21.8

Table 3. Percentage of Pages Containing Non-Textual Material

Journal	Total No. of Pages	Pages with Non-Textual Material No.	% of Total
BSL	220	138	63%
CJL	143	105	73%
FL	444	288	65%
IJAL	356	194	54%
JL	273	123	45%
Language	998	677	68%
L&S	282	201	71%
Lingua	448	243	54%
Linguistics	264	179	68%
CAnthr	524	232	44%
CLTA	269	213	79%
JVLVB	987	670	68%
MLJ	766	383	50%
MT	84	57	68%
Phonetica	252	156	62%

Table 4. Results of the Special-Symbol Count

Journal	Language(s)	No. of Pages Examined	Range in No. Per Page	Avg. No. Per Page
BSL	French	22	19-139	95
CLTA	French (1 article in English)	26	0-175	73
IJAL	English	33	0-293	50
Lingua	English French German	44	0-233	27
Language	English	90	0-204	27
CJL	English French	13	0- 73	25
Phonetica	English French German	25	0- 74	21
Linguistics	English German	25	0-105	13
FL	English	42	0-102	12
JL	English (1 article in Russian*)	24	0-119	8
CAnthr	English	51	0- 65	6
MLJ	English (1 review article in Italian)	52	0- 53	6
L&S	English	27	0- 44	5
MT	English	8	0- 11	3
JVLVB	English	94	0- 49	3

*Printed in Cyrillic

REFERENCES

1. Griffith, Belver. A study of information patterns of the Linguistic Society of America membership. Washington, D.C.: Center for Applied Linguistics, forthcoming. NSF GN-771.

2. Rice, Frank; Guss, Allene, eds. Information sources in linguistics. Washington, D.C.: Center for Applied Linguistics, 1965, vii, 42 p.

3. Comite International Permanent des Linguistes. Linguistic bibliography for the years 1939-1947, 1948, [et seq.] and supplement for previous years. Utrecht/Antwerp: Spectrum, annual.

4. REQUIREMENTS OF THE CHARACTER SET: A SAMPLE OF LITERATURE IN COMPUTATIONAL LINGUISTICS

4.1. Purpose

System data banking based on bibliographic input and computer-aided publication of a monthly bibliography were to be started in a pilot project whose first results are described in section 5. For computer-based publication of bibliographic data in the field of computational linguistics, the following information was sought. How many and which alphabetic symbols were utilized in the current bibliographical literature of the field and in references in its representative primary publications? How often did non-roman alphabet and other special symbols occur? Which transliteration systems were most commonly used? If priorities concerning characters to be developed had to be established, what was the distribution by language of the literature covered by major bibliographies in the field? What were the languages of tools used by researchers in the field, as seen from the references in primary works and periodicals?

4.2. Procedure

4.2.1. Selection of Sample

Some important bibliographies available in the United States and either wholly or partly devoted to computational linguistics are listed in table 1 (see p. 46). All but the Bibliographie Critique de la Statistique Linguistique are serial bibliographies. Bibliographie Critique was included as a major compilation which appeared in several of the bibliographical lists from which the primary works to be analyzed were chosen.

The primary works in the sample were those English-language primary works (monographs and articles) listed in two or more of the references in the following survey articles or reports: Susumu Kuno and Anthony G. Oettinger, "Computational linguistics in a Ph.D. computer sciences program" (Communications of the ACM, 1968, 11:12), Winfred P. Lehmann, "Computational linguistics" (Seminar on computational linguistics, Washington, D.C., National Institutes of Health, 1966), Warren Plath, "Mathematical linguistics" (Trends in European and American linguistics 1930-1960, Spectrum, 1961), and Milka Ivic, "Mathematical linguistics" (Trends in linguistics, Mouton, 1965).

The authors of the primary works appear in table 4 (see p. 53). Appendix C is a bibliography for the works by these authors used in the study. (In this study, the term 'primary works' is used to mean only monographs and articles in monographs, but not periodicals which are treated as a separate category.)

The periodicals for the study were selected from those cited by the Association for Computational Linguistics (ACL) as "journals [which] regularly contain a substantial number of pertinent articles". Included in the sample were those journals in the ACL listing which also appeared in the list of periodicals in the 1966 or 1967 Rand Computational Linguistics Bibliography, excluding three Russian journals and a Czech journal which could not be obtained for the study. 1966 was the most recent year for which all issues of the journals selected were available. A reference count was made for only those articles covered in the 1966 or 1967 issues of the Rand Bibliography.

Table 1. Secondary Sources Scanned

Source	Abbreviation	Sections Scanned	Issues Scanned
Bibliographie Critique de la Statistique Linguistique	BC	All sections	Not a serial bibliography (1954)
Bulletin Signalétique 24	BS	VII b - Applications of information theory to linguistics	
		XII f1 - Theory and problems in literary translation	
		XII f3 - Automatic translation	Vol. 22, Nos. 1-4 (1968)
Computational Linguistics: Bibliography, 1967	CL	All sections	One issue annually (1967)
Computing Reviews	CR	3.42 - Language translation and linguistics	Vol. 9, Nos. 2-11 (1968)
The Finite String	FS	All sections	Vol. 4, Nos. 1-10 (1967)
Linguistic Bibliography	LB	Translation, Mathematical linguistics	Vol. 21 (1966)
1967 MLA International Bibliography	MLA	Mathematical and computational linguistics	Vol. 83, No. 3 of PMLA (1967)
[Bibliography] La Traduction Automatique	TA	All entries	Vol. 5, Nos. 1-4 (1964)

4.2.2. Method

All entries from the bibliographies in the sample and all references and footnotes in the primary works in the sample were coded by language on work sheets of which figure 1 is a typed example. (The originals were filled in by hand.)

Entries (=References)

Language	Reference	Element	Embedding	Transliteration*	Non-roman Script	Special Symbols
English	135					
French	4	2				
Russian	4				Cyrillic	
German	10	3				
Czech	2	2				
Hungarian						
Polish						
Rumanian						
Italian						
Other (Latin)	2	4				
Other (Danish)	1					
Other (Norwegian)		1				

*Board of Geographic Names, Library of Congress, Linguistic Bibliography, Rand, International Standards Organization, National Federation of Science Abstracting and Indexing Services.

Figure 1. Example of Work Sheet

Since portions of bibliographic citations might occur in different languages, the form was divided into 'Reference', 'Element', and 'Embedding'. An entry whose title was in German was coded as 'Reference' in German. Or, an entry like

> Blass, F. (ed). Hyperidis orationes sex cum ceterarum fragmentis: Index vocabulorum composuit H. Reinhold, 141-176. Leipzig: 1894

was coded as a reference in Latin. Several elements of an entry might appear in a language other than that of the title. There might be a foreign-author name, the item might be published in another country in which case source and publishing data might be in another language. In this study, an entry was defined as having a foreign 'Element' if the title of the source was in a language different from that of the title of the work itself. An entry like

> Ajdukiewicz, K. Die syntaktische Konnexitat. Studia philosophica, 1935, 1, 1027

was therefore coded as a German 'Reference' with a Latin 'Element'. Defined and extracted as 'Embeddings' were those symbols, words, or phrases relatively frequently 'embedded' in titles of linguistic works to illustrate sound- or grammatical phenomena in another language. Thus, an entry like

> Lynch, Irina. Russian -СЯ verbs, impersonally used verbs, and subject/object ambiguities. In: 1961 International conference on machine translation of language and applied language analysis (London: 1962), p. 477.

was coded as an English 'Reference' with a Cyrillic 'Embedding'.

A notation was made of whether an entry in a non-roman alphabet language was printed in original script or transliterated. If the latter, an attempt was made to determine the transliteration scheme employed. Mathematical and logical signs were counted as 'Special Symbols'. The numbers in figure 1 represent the totals for a particular work.

4.3. Results

The results of the count are given in tables 2, 3, and 4. In all three tables, the figures for each language may be composite, the portion preceding the first period representing full references in that language, the portion after the first period representing elements, and the third portion representing embeddings.[1] Only full references are considered in the count of totals and the determination of percentages. In the tables for primary works and periodicals, citations printed in Russian Cyrillic are indicated by an asterisk, figures without asterisk represent transliterated Russian entries. All other non-roman alphabet languages occurred only in transliterated form with the exception of embedded Greek symbols. The journals Computational Linguistics and Prague Bulletin of Mathematical Linguistics were not printed in hot type.

[1]The composition of the total of 'Other' languages found in references, elements, or embeddings is as follows. Table 2: Arabic 1, Bulgarian 8, Chinese .03, Dutch 13, Finnish 1, Icelandic 1, Japanese 1, Korean .01, Latin 45.16.22, Latvian 1, Slovak 2, Spanish 4.02.1, Ukrainian .03. Table 3: Czech 9, Danish .01, Dutch .01, Hungarian 9, Latin 6.01, Norwegian .01, Sanskrit 1, Swedish 1. Table 4: Chinese 1.01.1, Czech 6.05, Danish 3, Dutch 2, Greek .01, Hungarian 2.01, Italian 12.06.01, Japanese 2, Latin 9.20, Norwegian .01, Polish 2, Romanian 3.01, Serbo-Croatian 1, Spanish 2.03, Swedish .08, Welsh 1.

Table 2. References in Secondary Sources by Language

	BC	BS	CL	CR	FS	LB	MLA	TA	Total	%
English	957.08.4	8	466.48.1	125	626.31	107.23	41	38	2,368	65.5
French	204.61.6	15	54.32	.01	58.30	37.18	7	30.01		11.5
Russian	5	.03.1	108.10		173.09.1	35.04	3	.14	324	9.0
German	183.07.2	2.02	16.09.1	.29	19.09	31.05	2	2.03	255	7.0
Czech	23.02	.02	11.05	1	10.02	35.06	1		81	2.2
Hungarian	6.01	.04	3.01		4			.02	13	.3
Polish	9.04	.02			1	18			28	.8
Rumanian	1.01				2	21	1		25	.7
Swedish	10								10	.3
Danish	4			1					5	.1
Norwegian	2								2	.1
Italian	5		2		3	2.01	3		15	.4
Other	58.10.23	.01	4.01		2.01	13.05		.07	77	2.1
Symbols	Greek		Greek	Greek	Greek					
Total References	1,467	25	664	127		299	58	70	3,618	100

Table 3. References in Periodicals by Language

Year and Number of Articles	English	French	Russian	German	Italian	Rumanian	Other	Total References
American Documentation 1966 [18]	116	1	12.04	2			2	133
Behavioral Science 1966 [4]	135.02	3	12.16	2			.01	152
Beiträge zur Linguistik und Informationsverarbeitung 1966 [8]	4			25				29
Communications of the ACM 1966 [39]	219.01	9	1	7.01	2			238
Computational Linguistics 1966 [4]	15	2	4*	6			14	41
Computers and the Humanities 1966 [1]	2							2
Foundations of Language 1966 [10]	169	3		17.01	1		.01	190
Information and Control 1966 [8]	82	1	.01					83
Information Storage and Retrieval 1966 [2]	22							22
Journal of Chemical Documentation 1966 [15]	68							68
Journal of Linguistics 1966 [8]	100	9	1	5	3	5	.02	123

Table 3, cont.

Year and Number of Articles	English	French	Russian	German	Italian	Rumanian	Other	Total References
Journal of the ACM 1966 [12]	96.01	1	1.03	3.04				101
Journal of Verbal Learning and Verbal Behavior 1966 [20]	156	3	1	1				161
Language 1966 [3]	75.01	2					6	83
Linguistics 1966 [4]	11.3	15	7	1			1	35
La Linguistique 1966 [2]	2	9.01	5	3				19
Mechanical Translation 1966 [11]	73	6						79
The Prague Bulletin of Mathematical Linguistics 1966 [4]	4		18*	15			2	39
Statistical Methods in Linguistics[1] 1966 [4]	24	.01	5*	2		4		35
TA Informations 1966 [3]	20.01	19.02	7.02	1		1		48
Word 1966 [4]	12	1		2			1	16
Total References	1,405	84	74	92	6	10	26	1,697
Percent	82.80	4.95	4.35	5.45	0.30	0.60	1.55	100

[1]Both Cyrillic and transliterated references given.

Table 4. References in Primary Works by Language

Work(s) by	English	French	Russian	German	Other	Total References
Abernathy	5.02		139			144
Akhmanova et al.	112.05	14	55	2.02	2.05.1	185
American Mathematical Society	197	9.04	6	7.01	3.03	222
Bobrow et al.	121		.02		.03	121
Borko	456	1.02		1	1.04	459
Carnap	3					3
Carroll	19			1		20
Chomsky	194	6.01	2.02	3.01	1.03	206
Delavenay	19	6	5			30
Edmundson	106.04		11.02	6.01	.01	123
Garvin	308.01	12.03	.01	9	1.06	330
Garvin and Spolsky	96			2.01	.01	98
Harper	33.07		16.11			49
Hays	194.01	5	1.02			200
Herdan[1]	336	15.02	7*	51.03	24.11	433
Hockett	63.01	5.01		2	1	71
Ivic	53	6	19	2	1	81
Kiefer	50.01	10.02	17*.02	5	4.01	86
Kuno	54	1		.01		55
Lehmann	16.01	1.01		1	1	19
Locke and Booth	89.02	2		3	1	95
National Physical Lab	185.01	13	21.01.2*			219
Oettinger	197	8	15.01	3		223
Papp	7		511*.01	3.01.1	3.04	524
Plath	83	5.01	1*	8	2.03	99
Salton	132.01	10.06			.02.1	142
Simmons	116					116
Whatmough	16	3	1		1	21
Yngve	8	.01		1		9
Total References	3,268	132	827	110	46	4,383
Percent	74.55	3.05	18.85	2.50	1.05	100

[1]Three references in Cyrillic, four transliterated.

4.4. Remarks on Problems

An important objective of this study, as stated above, was to determine which transliteration methods were most commonly used by language sciences scholars and by the editors of language sciences journals and bibliographies. The data collected for the bibliographic citation study proved unusable in this connection because the textual information contained in the references was insufficient to permit accurate determination of the transliteration employed. If there were establishable and consistent preferences among linguists for certain transliterations, a different approach would be necessary to determine them. It was sometimes difficult to identify some Latin-alphabet languages in bibliographic references. To determine conclusively, for example, whether an entry was in Faroese or Frisian would involve considerable effort. The gain of discovering a letter or diacritic not already contained in the alphabets of languages used in other references had to be weighed against the labor required to make a positive identification. It was usually considered more efficient to rely on an informed guess which took into account available publishing data.

4.5. Conclusions

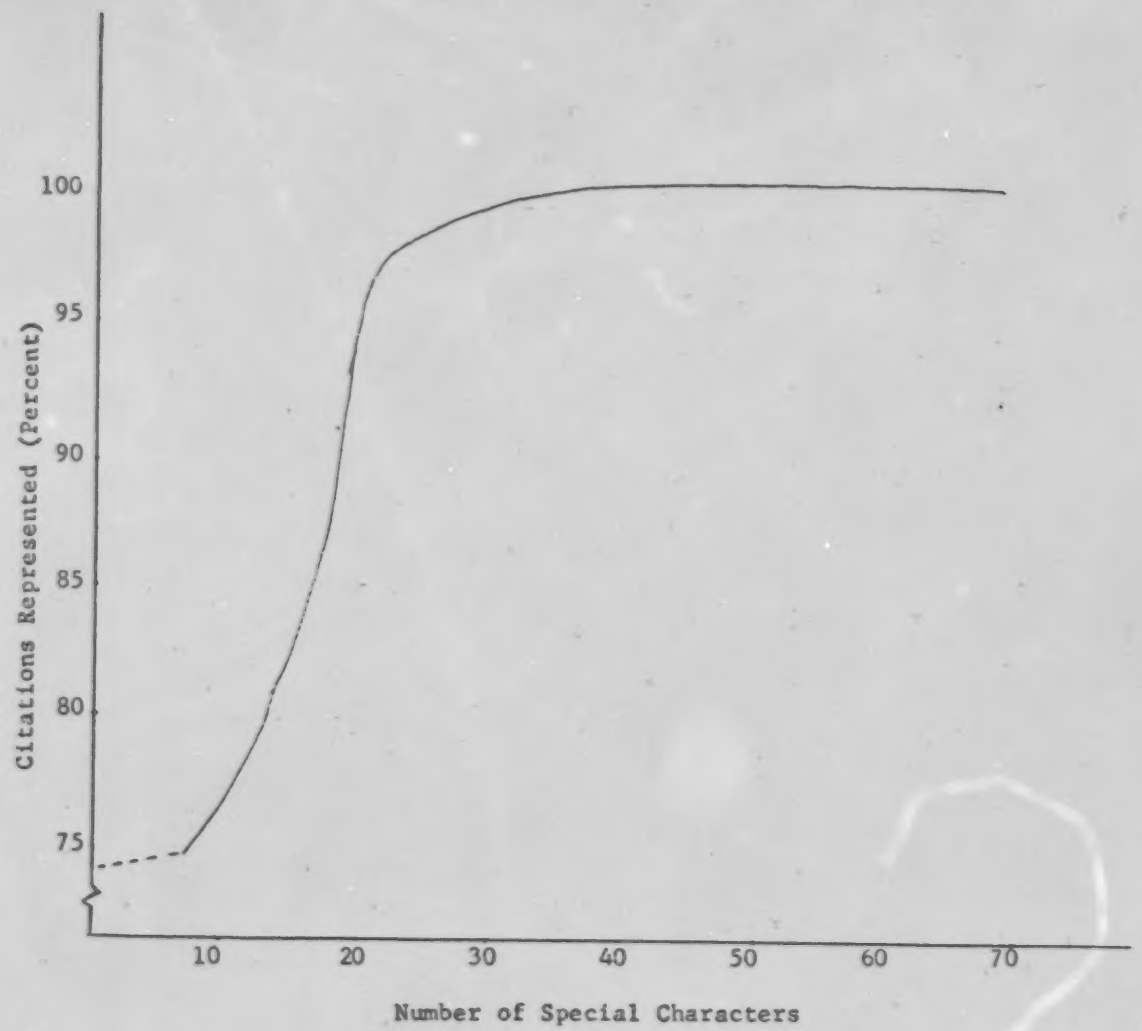
Analysis of this sample of secondary publications and references in primary publications in a subfield of linguistics led to the following conclusions:

A desirable character set for quality secondary publications should accommodate the letters (alphabetic or transliterated) of a fairly wide range of languages. In the bibliographies examined, full bibliographic citations were found in 22 languages. To represent these languages, approximately 70 roman-alphabet letters and modifications would be needed in addition to the English alphabet. Figure 2 (see p. 56) shows that slightly fewer than 75% of bibliographic citations in computational linguistics literature can be printed with the letters of the English alphabet. The requirement for additional roman-alphabet characters rises steeply to about 97% when citations in French, Russian, and German are added to those in English. The cumulative curve then levels off into a long tail representing the additional character requirements for eighteen other languages. These facts suggest that the need for expanded character set capabilities in language sciences publications is relatively greater than that in other disciplines. Hefner et al. report, for example, that about 95% of the total literature covered by Chemical Abstracts can be printed without special characters [1].

Figure 2 represents the averaged cumulative special-character requirement for the three types of literature used in this study. The distribution within each category (secondary sources, periodicals, primary works) can be seen from tables 2, 3, and 4. Both primary works and bibliographies employed a wide range of languages in references. In periodical articles, however, the number of languages in which references were given was much lower than that for bibliographies. Guttman [2] arrived at a similar conclusion with respect to social sciences periodicals, noting that very few foreign-language tools appeared to be used by social scientists in their current research as reported in the journal literature. On the other hand, while the range of languages employed in the primary and secondary works in computational linguistics was wide compared to that in periodicals, the percentage of references in languages other than English, French, Russian, and German was low, only 1.05% in the primary works, 2.45% in the periodicals, and 7% in the bibliographies.

References in both primary works and in journals were not infrequently given in Cyrillic. In the secondary sources no oriental or other non-roman scripts were found except for Greek symbols of which there were 26 including some upper case ones. This would appear to be negligible in a sample of 3,618 citations. Greek-alphabet letters (without accents) are available on many phototypesetters though Cyrillic ones are not.

Generalizations to the wider field of the language sciences literature are not easy to make. The sample studied was taken from a subfield of linguistics not as centrally concerned with language description as other specialties in linguistics. A projection of character set needs set needs based on this study would presumably be too low. It is clear from this study that the development of an adequate set of special characters for even a limited subfield of linguistics is urgent and will be costly.



ā ą ă á å æ ć ď đ ḍ ǧ ě ę ġ ḥ
ī í ì ķ ļ ł ń ņ ñ ń ň ó ò ø ő ŕ
ř ş ś ș ť ť ț ț ú ū ů ű ý ź ź ż þ

ä ö

č ė ı š ž

à â ç è é ê ë î
ï ô œ ù û ü

Figure 2. Number of Unique Special Characters Required to Print a Given Percentage of Bibliographic References in the Primary and Secondary Literature in Computational Linguistics

REFERENCES

1. Hefner, R. G.; Keesecker, P. M., Rule, D. F. Keyboarding chemical information. Journal of Chemical Documentation, 1967, 7:4.232-236.

2. Guttsman, W. L. The literature of the social sciences and provision for research in them. Journal of Documentation, 1966, 22:3.186-194.

5. THE LANGUAGE AND AUTOMATION BIBLIOGRAPHIC PROJECT

This section describes some of the steps taken by the LINCS pilot project in the experimental development of an automated bibliographic publication service. These efforts included a prototype international network of cooperating centers supplying information for a bibliographic data base covering all current research literature in the language sciences and related fields.

The steps described include the generation of bibliographic data, their collection, editing, tagging for machine identification, input, extraction and preparation for printed display. At each step, different people -- bibliographers, abstractors, editors, keyboarders, computer personnel, and graphic arts designers -- were involved. The LINCS project was guided by the philosophy that each of the component stages should be kept distinct to the extent possible in order to allow those involved at each stage to operate in their customary ways. At the same time, the steps were to be combined in a single, uninterrupted operation from citation and abstract generation to typeset display with no duplicate transcription or keyboarding of data and with a minimum of tagging and editing.

In the first stage, the small bibliography, 'Recent Publications in Computational Linguistics' which appears in the Association for Computational Linguistics' monthly newsletter, The Finite String, was photocomposed. In connection with this bibliography, the project began its first serious exploration of both phototypesetting techniques and possibilities and of input procedures leading to a mechanized data base and publication capability.

Experiments were conducted with several United States firms using different phototypesetting equipment. Given the goal of developing a fully automated data base from which different kinds of publication outputs could be obtained, it seemed desirable for the project staff to acquire proficiency in in-house coding of textual and bibliographic material according to a given firm's typesetting system. As a preliminary, the project experimented with a special data form to be completed by the cooperating bibliographers and to be directly keyboarded without intervening tagging, rearrangement, or transfer of the data. This initial data form was designed to cover all types of publications, books as well as technical reports, journal articles, etc. The single form did not, however, easily accommodate all the descriptive elements ideally collected for diverse document types. It also tended to confuse the bibliographer because, depending on the document being cataloged, certain data fields had to be completed and others left blank. In addition, the ordering of bibliographic elements in the entry as obtained from the bibliographer was different from that in the printed output and thus required further editing.

In the second stage, during which the publication of Language and Automation was begun, input procedures were redesigned to handle the significantly greater volume of bibliographic data. A comprehensive, flexible, and structured definition of the data set was evolved which, in terms of the elements included, was compatible with that of the other bibliographic services. Key specialists in bibliographic computer systems and photocomposition provided continuing consultation as the system developed, including the programming for the automated file maintenance and publication system.

Prior to the redesigning of the LINCS data form the project contacted many organizations which had amassed experience in dealing with the problems of compatibility and standardization in information interchange. The combined single input form was abandoned in favor of a separate form for each major document type for several reasons. (1) In the LINCS cooperative network source data acquisition, i.e. cataloging and abstracting, was performed in several countries. Standardized forms made it easier for the different bibliographers to supply all the data required in the LINCS system in a uniform manner. (2) Different forms for different document types simplified the bibliographer's task. Each form was tailored to a given document type; each elicited only those data elements required for bibliographic description of the given document. (3) The editor's task was thus simplified as well. A bibliographic entry uniquely identified was more easily evaluated as to completeness and accuracy. (4) By far the greatest advantage derived from the fact that certain of the bibliographer's, editor's, and keyboarder's functions were combined in one operation. The intermediate step (i.e. tagging) between the bibliographer and the keyboarder was eliminated by having the tags required for computer identification of data elements on the forms so that keyboarding could be done directly from these. (5) Explicit identification of all functional data elements (e.g. author, title) was essential for development and maintenance of a large bibliographic data base. This explicit identification was also the first step in an automated publication system. A file management program could rearrange and extract from the data base portions for publication, and a typesetting program could translate the functional identification codes into typographic specifications for the desired printed output.

Eight data forms were developed for Language and Automation and sent to the LINCS bibliographers. They were modeled to some extent on the bibliographic data information sheets used by the Chemical Abstracts Service. The color-coded forms covered the following document categories: Book (canary yellow); Item from a Serial (goldenrod); Item from a Collection (green); Technical Report (buff); Dissertation (pink); Translation and Conference Proceedings (white); Review (blue); Serial Announcement (white).

The flow of bibliographic data in the LINCS publication program, from editing and keyboarding of the filled-in data form to the typeset output is outlined in the simplified flowchart in figure 1. The boxes labeled 1-4, the first three of which represent the bulk of the involvement of the LINCS editorial staff, are further explained in the text following figure 1, and their 'products' are exemplified in figures 2-6.

Figure 2 (see box (1) in figure 1) is a sample filled-in data form (for a journal article). The far left column of the printed form identifies the data field for the bibliographer. The second column contains the codes which tag the data fields for computer manipulation. The bibliographic data structure is so designed that, in addition to explicit identification, implicit identification may be made on the basis of the order of elements within a given information field. Such elements are printed on the form when they correspond to conventional information units -- series, year, volume, etc. under 'Pagination Grouping'. Otherwise, the bibliographer and/or keyboarder is instructed to type them in a specified order. Under 'Corporate Name', for example, the following units of the corporate name are identifiable:

First position:	Largest unit
Second position:	City of smallest unit
Third position:	State of smallest unit
Fourth position:	Country of smallest unit
Fifth position:	Smallest unit
Sixth position:	Acronym largest unit
Seventh position:	Acronym smallest unit.

Implicit tagging provides a powerful search and retrieval capability without making the keyboarding cumbersome through either a fixed-format approach or excessive use of tags. It is necessary only to delimit each position and to designate (by 00, for example) empty positions if they occur at the beginning or within a sequence of positions making up an information field. Since LINCS bibliographic services requirements were initially restricted to typeset display in Language and Automation, the implicit tagging and search capability had not been utilized at the time of writing. Both the automated publication capability and the limited file management program being designed by LINCS are entirely flexible, however, and allow for expansion and modification as requirements change and/or the financial resources for full operational development and refinement of LINCS bibliographic services become available.

Figure 3, (see box (2) in figure 1) shows the data from the sample form after keyboarding on the IBM Magnetic Tape Selectric Typewriter (MTST). The keyboarding is done directly from the forms in free format. The publication type (i.e. book, technical report, etc.), accession number, author and subject category form the header line for sort purposes. The information in each distinct data field is then typed on a separate line with its respective code. The MTST cartridge is then converted to magnetic tape.

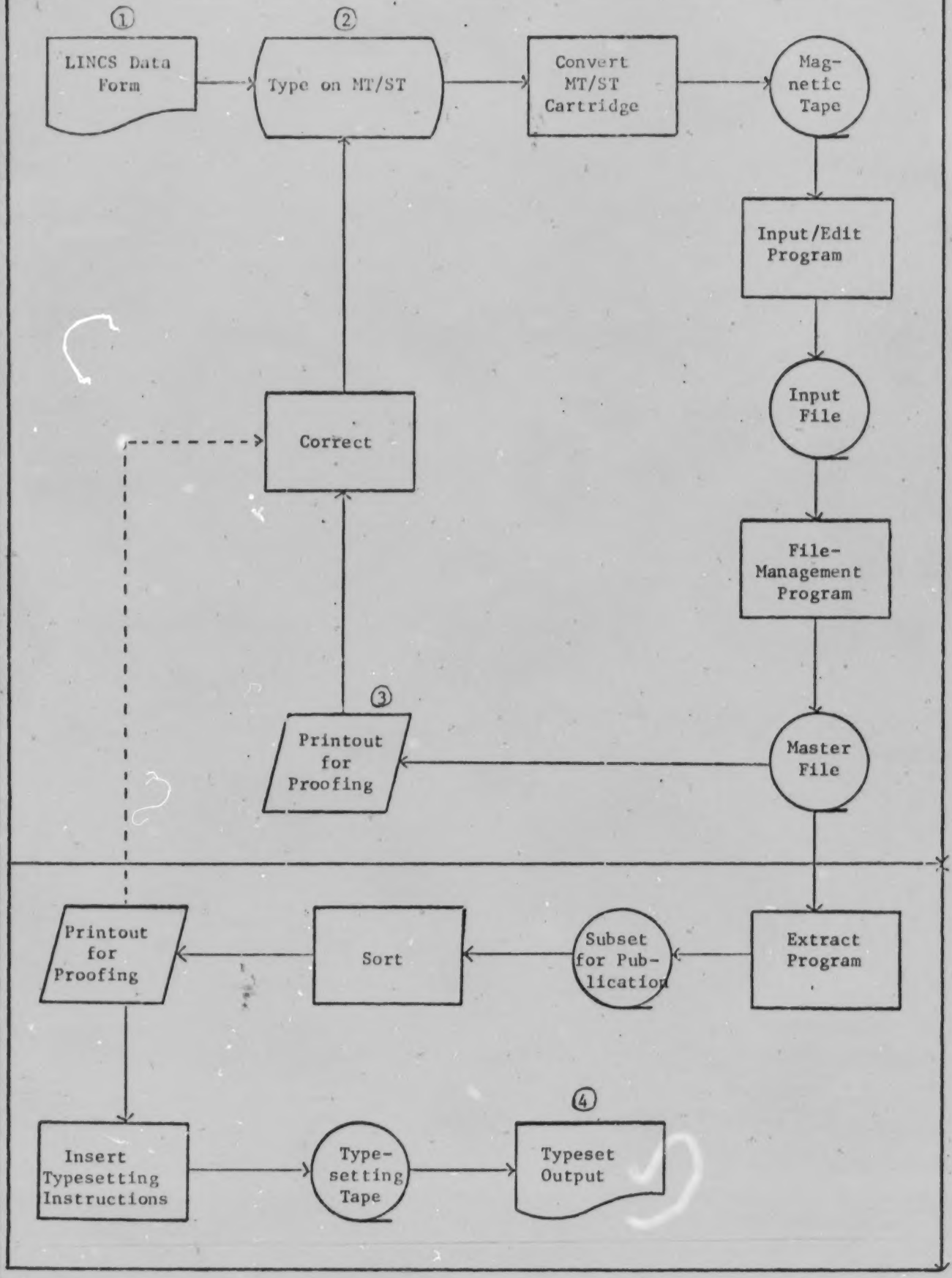


Figure 1. Bibliographic Data Flow in LINCS Pilot Publication Program

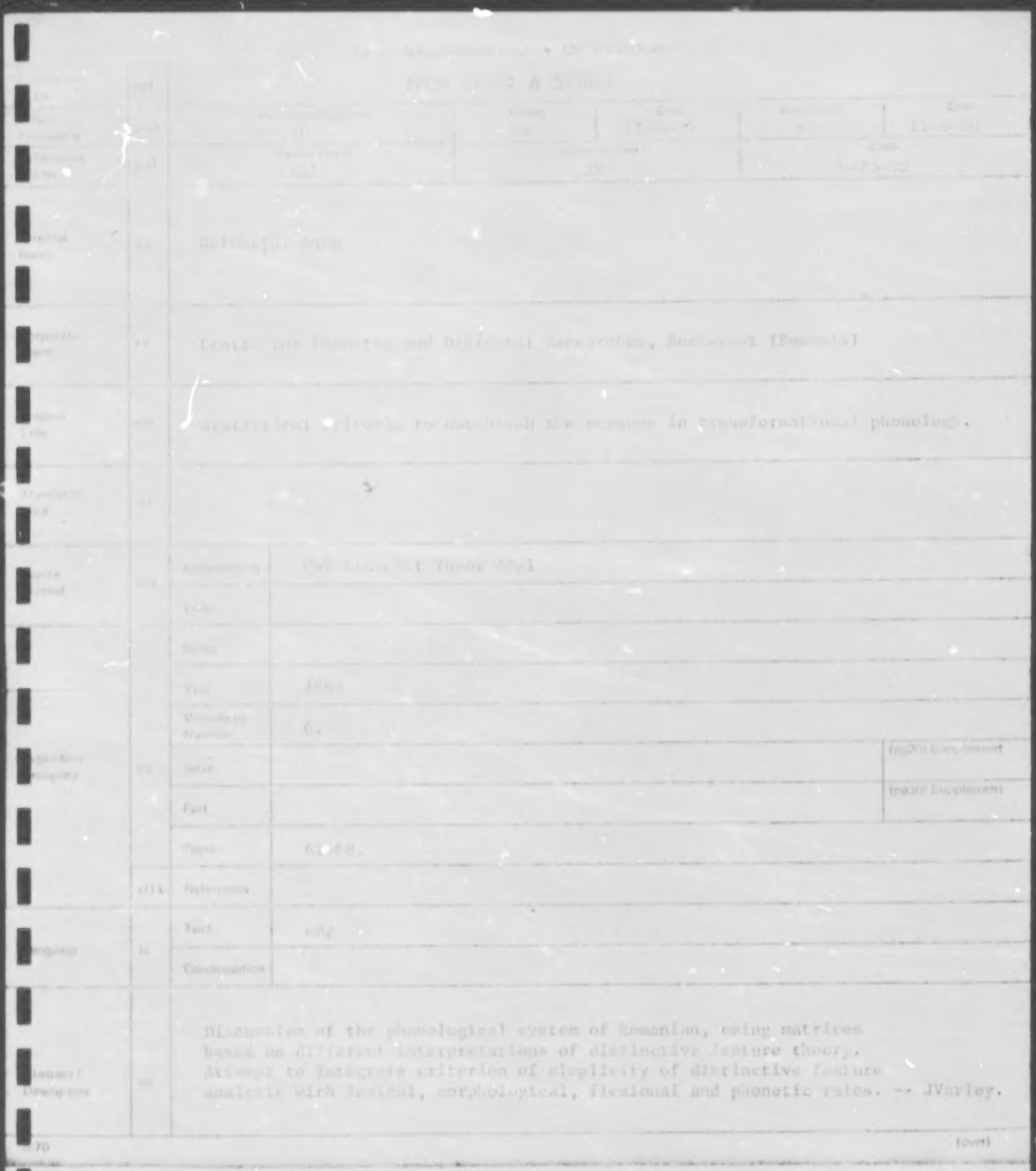


Figure 2. Sample Completed Data Form. (On reverse side of the form there is space for the subject classification of the item.)

```
+ty1,396 BELCHITA ANCA,sc000QL,sc001LA sc010QL,sc011LA
+pd0,896,be,1-19-71,bz,1-19-71
+pd1,cal,jv,9-23-70
+pn,Belchita, Anca
+cn,Centre for Phonetic and Dialectal Researches, Bucharest (Romania)
+ti0,Statistical criteria to establish the economy in transformational
phonology
+sol,Cah Linguist Theor Appl
+pg,1969, 6.61-68
+ab,Discussion of the phonological system of Romanian, using matrices
based on different interpretations of distinctive feature theory.
Attempt to integrate criterion of simplicity of distinctive feature
analysis with lexical, morphological, flexional and phonetic rules
```

Figure 3. Entry Typed on MTST for Input to LINCS Data Base.

Figure 4, (see box (3) in figure 1) is a sample printout for the entry. After proofing, corrections, keyed to the accession number for the entry, are typed on the MTST. They are input into the system routinely, together with new entries. A second printout is obtained when all entries marked for announcement in Language and Automation (or any other LINCS publication) have been extracted from the general data base and sorted under their respective subject category headings.

Figure 5 shows mock-up of a third printout which is not normally obtained. It is included to exemplify the conversion from functional bibliographic codes to typesetting instructions (including assignment of publication number and punctuation) which is effected by the publication program. The typesetting specifications are expressed as coded commands for which macros (the # or ½ in figure 5) defined in a look-up table can be used. The typesetting specifications shown are those used by LINCS in the photocomposition of the early experimental issues of Language and Automation. The LINCS publication program is, however, completely flexible and independent of a specific typesetting system or 'language'. Following this conversion the tape with complete typesetting instructions is sent to the phototypesetting firm.

ACCESSION NUMBER: 000896
OL PUBLICATION NUMBER: 7100140

ITEM CODE	ITEM VALUE
ty1	BELCHITA ANCA sc0000L; sc0001LA; sc0100L; sc0111LA•
pd0	896,be,1-19-71,bz,1-19-71•
pd1	cal,jv,9-23-70•
pn	Belchita, Anca•
cn	Centre for Phonetic and Dialectal Researches, Bucharest (Romania)•
ti0	Statistical criteria to estatlish the economy in transformational phonology•
so1	Cah Linguist Theor Appl•
pg	1969, 6.61-68•
ab	Discussion of the phonological system of Romanian, using matrices based on different interpretatio of distinctive feature theory. Attempt to integrate criterion of simplicity of distinctive featu analysis with lexical, morphological, flexional and phonetic rules•

Figure 4. Sample of Printout for Proofreading after Item has Passed through Input/Edit Stage.

896

#Pelchita, Anca•

%Centre for Phonetic and Dialectal Researches, Bucharest (Romania)•

½Statistical criteria to establish the economy in transformational phonology•

¼Cah Linguist Theor Appl•

1969, 6.61-68•

@Discussion of the phonological system of Romanian, using matrices based on different interpretations of distinctive feature theory. Attempt to integrate criterion of simplicity of distinctive feature analysis with lexical, morphological, flexional and phonetic rules• *

Figure 5. Mock-up of Printout from Publications Subfile; Tape Ready for Typesetter.
(#, %, ½, ¼, and @ represent typesetting commands defined in a look-up table.)

Figure 6 (see box (4) in figure 1) represents the final entry display, typeset in the format of Language and Automation. The figure exemplifies the most intractable problem LINCS faced in the production of Language and Automation thus far, namely, the unavailability of special characters in most existing automatic typesetting systems. Diacritics had to be inserted manually in Language and Automation. Recent developments in the phototypesetting industry seem to indicate that at least a partial solution to the LINCS character set problem is technically feasible now and may be economically feasible in the near future.

896.70
BELCHIĂ, ANCA
Centre for Phonetic and Dialectal Researches, Bucharest (Romania)
Statistical criteria to establish the economy in transformational phonology. Cah Linguist Théor Appl. 1969, 6.61-68.

Discussion of the phonological system of Romanian, using matrices based on different interpretations of distinctive feature theory. Attempt to integrate criterion of simplicity of distinctive feature analysis with lexical, morphological, flexional and phonetic rules.

Figure 6. The Typeset Entry in the Final Output (Reproduced with the permission of Language and Automation.)

6. Conclusions

The most significant conclusions resulting from this study of graphic and bibliographic characteristics of current research literature in the language sciences may be summarized as follows:

(1) Bibliographic services in the language sciences, in the aggregate, show most of the data elements which current documentation standards and groups concerned with bibliographic information interchange consider essential in the description of printed records.

(2) Bibliographies differ widely in the range of document types covered, languages employed in citations, and typographic approach to representation.

(3) In core linguistic journals, approximately 23% of the contents consist of tables, graphs, tree structures, and other non-textual materials, an average of 62% of all pages contain such material; and 25 special symbols are found on an average page.

(4) In a sample of literature in computational linguistics, 75% of bibliographic citations were in English, 97% were printed in the four major languages, English, Russian, French, German; some 70 non-English alphabetic symbols are required to print all citations.

Two somewhat speculative measures of the relative importance of different types of publications were obtained. (1) The 12,893 citations in the 1966 issue (the latest issue available at the time of the study) of the international Linguistic Bibliography were classified into the nine categories shown in figure 1 (see p. 70). Journal articles comprised 41% of all entries. The second most important category, book reviews, also made up a substantial portion of the entries in other linguistic bibliographies. The lack of a comprehensive and timely review or abstracting service covering the entire literature of the profession perhaps accounts for the prominence of review citations in annual bibliographies. Also, about one fourth of the annual volume of several major linguistics journals is devoted to reviews of new publications. (2) An examination of the bibliographic citations in nine 1967 core linguistics journals (table 1, see p. 71) indicated that non-periodical literature had the highest priority with journal articles taking second place. This result was in agreement with similar analysis by Guttsman (1966) which he interpreted as an indication of the significance of different document types as research tools in social scientists' work.

A very tentative estimation may be made concerning the probable magnitude of the character set requirements in a language sciences information system designed to cover the relevant literature and provide services including both secondary and primary publications of high typographic quality.

Linguistic materials are printed in many languages and in several alphabets. A large symbol set would be necessary even for secondary coverage of such materials. In the 1966 Linguistic Bibliography, citations were printed in fifty languages, 37 of which were based on Latin script, and 13 of which were based on various non-Latin scripts and transliterated. These alphabets and transliterated equivalents use some 220 unique letters (figure 2, see p. 73). For Greek the Linguistic Bibliography and other linguistics writings use the extended alphabet with accent marks, with a symbol set exceeding a hundred. Considering upper- and lowercase variants, the alphabetic letter portion of a basic linguistics character set would number approximately 650.

A second major set of graphics are the phonetic symbols linguists employ to describe the sounds of languages. In the secondary literature (bibliographies), these occur typically as "embeddings" in references to extinct languages or languages with a non-roman alphabet for which no generally accepted transliteration exist, or in citations containing in the title a sound or set of sounds dealt with in the text proper. While linguists of different schools may use somewhat different notational systems, the most important symbols are those of the International Phonetic Association (IPA). The IPA's basic vowel and consonant letters can be augmented by a number of other, discrete symbols or by digraphs to represent sound gradations unique to certain languages by diacritical marks to depict aspects of sounds having to do with point and manner of articulation; and by marks to represent length, stress, and pitch. Table 2 (see p. 72) summarizes basic IPA graphics, excluding those of its letter symbols which occur in the set of alphabetic characters in figure 2. It shows 120 unique phonetic symbols which, if added to the previous 650 symbols, would enlarge the linguistics character set to 770 (phonetic symbols do not have upper case variants). There are, furthermore, 22 diacritical tone and stress marks which occur over or beneath many IPA letters. The number of letters with which they can combine varies widely but the possible combinations number several hundred at least.

The phonetic symbols are more important in the primary linguistics literature than they are in bibliographies. In major monographic treatises and in core journals devoted to synchronic language description, the historical study of language, dialectology, or contrastive studies, the phonetics symbol set is entirely elastic, growing in size with the narrowness of description and specificity of sound representation which the linguist wishes to achieve. If, in a transcription of any language, a given sound or shade of sound cannot be represented by standard phonetic symbols, the linguist may use almost any number of ad hoc modifications. Also, an almost unlimited number of other script symbols, from non-roman alphabetic latters to logographs and hieroglyphics, are found in linguistics texts. For the primary linguistics literature, therefore, the character set may have to be considered essentially infinite. (Figure 3 (see p. 74) illustrates the use of special alphabetic and phonetic symbols in secondary and primary linguistics literature.)

Russian Cyrillic and Hebrew script, though not used as citation media in the Linguistic Bibliography, do occur frequently in other secondary publications. The approximately 110 combined characters of Hebrew and Cyrillic may therefore have to be added to the basic alphabetic and phonetic symbols.

Another set of symbols must be considered though it is not necessarily restricted to the linguistics literature. These are the 'global morphemes' -- numbers, mathematical and logical operators, and other signs used in scientific description -- and punctuation marks. They probably number at least two hundred.

A minimal character set for language sciences publications is therefore estimated at this time to comprise some 1,500 unique characters (including upper- and lowercase variants). This is of course not a workable set in terms of typographic output for which at least some diversification through different type faces is indispensable. Obviously, the design of an adequate character set for computer based photocomposing of language sciences publications will be a costly problem even if the "ideal" solution of a unique graphic for each symbol is abandoned in favor of letter-diacritic combinations. Detailed and precise documentation of special character needs and use will be an immediate task within the LINCS study of characteristics of the language sciences literature.

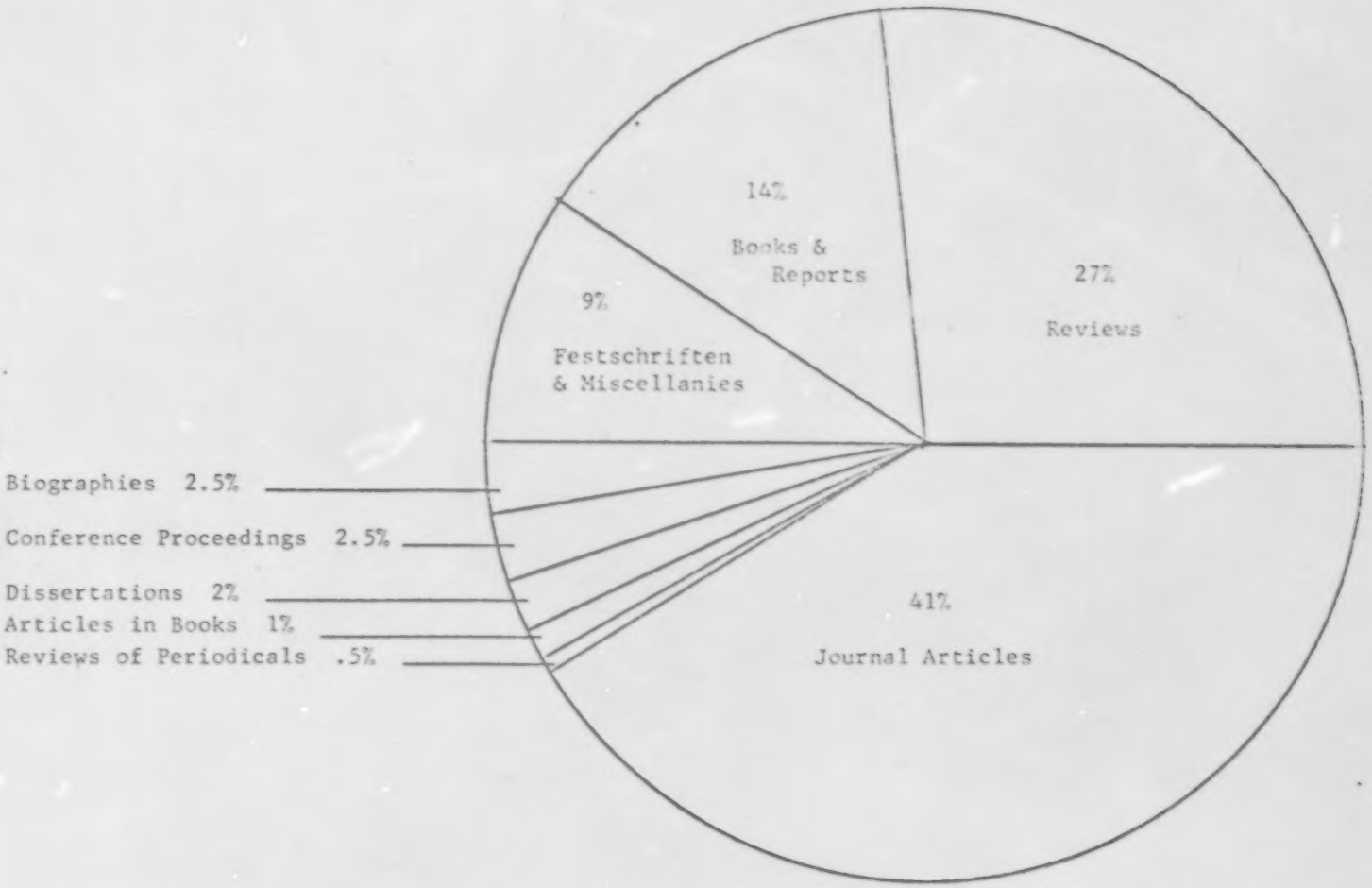


Figure 1. Distribution of Linguistic Literature by Type of Publication Outlet, Based on 1966 Linguistic Bibliography

Table 1. Analysis of References in Language Sciences Journals

Journal Examined	Number Articles	Type of Publication Cited					Unpublished Material	Number Citations
		Books & Reports	Miscellanies Festschriften Proceedings	Dissertations	Book Reviews	Journal Articles		
Bulletin de la Société de de Linguistique de Paris 62:1, 1967	8	58	14	1	-	65	-	138
Canadian Journal of Linguistics 12, 1966/67	10	38	4	3	-	21	1	67
Foundations of Language 3, 1967	11	84	31	1	2	56	5	179
International Journal of American Linguistics 33, 1967	21	82	8	3	1	90	15	199
Journal of Linguistics 3, 1967	9	79	20	2	1	48	10	160
Language 43, 1967	40	417	85	10	13	121	9	655
Language and Speech 10, 1967	21	87	36	7	1	127	5	263
Lingua 18, 1967	13	95	20	-	7	94	3	219
Linguistics 36 and 37, 1967	14	138	29	4	14	73	2	260
Total Citations 147 Articles		1,078	247	31	39	695	50	2,140
Percentages		50.4	11.5	1.5	1.8	32.5	2.3	100

Table 2. Letters of the International Phonetic Alphabet*

Vowel letters	
Distinctive symbols (normal alphabetic letters like e, ø, etc., omitted)	14
Consonant letters	
Distinctive symbols (normal alphabetic letters like w, v, θ, etc. omitted)	38
Other letters (special forms of consonant or vowel letters found in certain languages)	42
Diacritical marks in independent spatial position (the letter "o" is used as a substitute character):	
o˞ o˞ o˖ o˕ o˗ o˒ o˓ o˔ o˒ o˒ o˒ o˒ o˒ o˒ o˒ o˒ o˒ o˒ o˒ o˒ o˒ o˒ o˒ o˒ o˒	26
Total unique phonetic symbols	120

Diacritical marks in combination with letters:

o̰ o̤ o̬ o̤ o̥ o̜ o̪ o̺ o̻ o̼ o̽ o̯ o͡o

o̯ o̩ o̯ o̯ o̯ o̯ o̯ o̯ o̯ 22

*Extracted with permission from: The Principles of the International Phonetic Association (London: International Phonetic Association, University College, 1949, reprinted 1962), pp. 7-19.

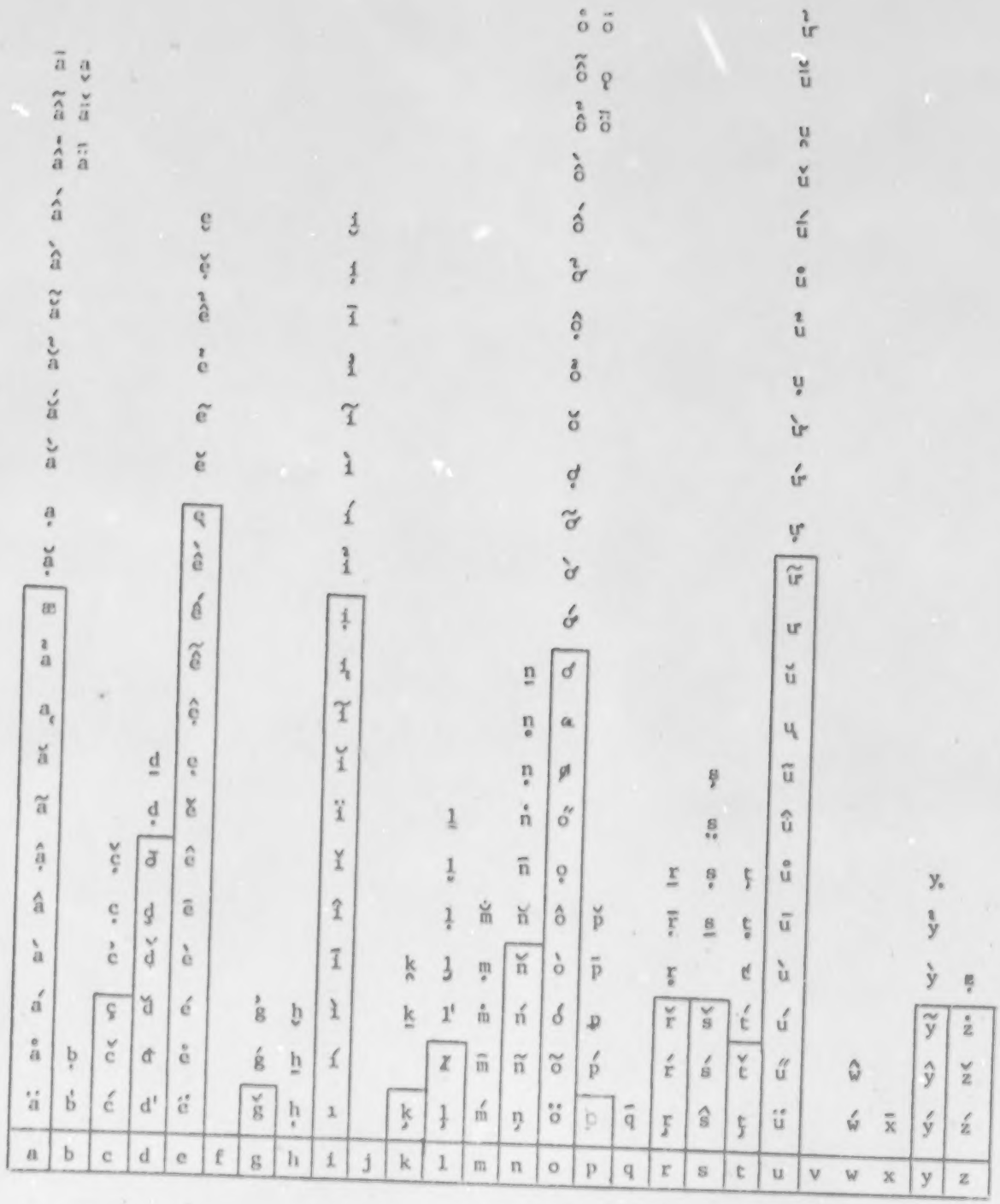


Figure 2. Character Set of Roman-Alphabet and Transliterated Languages Used in Citations of major Linguistic Bibliographies.

A. Vokale:

a (~ ɑ̯), ɑ̯, ȧ ~ ɑ̯, ɑ̯̇, ɛ, ɛ̇, ɛ̯, ɛ̯̇, ɛ̯̈, ɛ̯̇, ɛ̯̇, ɛ̯̇, i, i̯, i̯, i̯̇, i̯̇, i̯̇, ɪ, ɪ̯, ɪ̯̇, ɪ̯̇, ɪ̯̇, ɪ̯̇, ɔ̈, ɔ̯, ɔ̯̇, ȯ, ɤ̯, ɤ̯̇, ü; ȧ, ᵒ, ɑ̯, ɑ̯̇, ɑ̯̇, o, ɑ̯, ɑ̯̇ (~ ɑ̯̇), u̯, u̯̇, u̯̇, u, u̯; u̯̇, u̯̇; v, v̇, ᵊ, ɔ, ɔ̇, ɔ̯, ɔ̯̇, ɪ; v̇, v̇, ɔ̇, ɔ̯̇, ɔ̯̇, ɪ̯̇.

B Konsonanten:

ʾ (~ ʾ), k̯, k̯̇, g, g̯ (~ g), k̯, k, c, g, k̯, k̯̇, k̯̇, k̯̇ (~ k̯ʰ), t̯, ṫ̯, ṫ̯, t̯, t, ɒ, d, t̯, d̯, l, ɓ d̯, p̯, p, ʙ, b, p̈, ṗ̯, b̈; h̯, ɦ, h, ʜ, ʜ̈, χ, ɣ, ᵡ, ᵟ, δ, φ, β, β̈; j̯, j̯̇, j, ɥ, ɰ, ʏ ɕ, ʃ, ʃ̯, ʃ̯̇, ʒ, ʒ̯, s, z, z̯, z̯, ʂ, ʂ̯, ʃ̯, ś, ʃ̯̇, ź, ʒ̯, ʒ̯̇, ɾ, r, ɾ̈; ɹ, l̯, ɻ, l, l̯, ʟ, l̈; ŋ, ŋ̇, ɲ, n ɲ̯, ṅ, m, ṁ

§ 6. Im Kamassischen kommen folgende Diphthonge vor:

aį̯, ȧį̯, ɑ̯į̯, ɒį̯, ɒ̯į̯, ɒ̯į̯, ɛį̯, ɛ̯į̯, ɛ̯į̯, ɔį̯, ᵊį̯, ɔ̯į̯, ɔ̯̇į̯, ɔ̯̇į̯, ᵊį̯, oį̯, o̯į̯, ȯ̯į̯, uį̯, u̯į̯, u̯̇į̯, uį̯, ʷį̯, ia, iȧ, ʲȧ, iɪ, iɛ, ɪɛ, ɪɛ, ɪɛ̯, ʷɛ̯, ʷɔ̯, ʷɔ̯̇, au̯, ʲu.

	Singular	Dual	Plural
1. P.	$ku^{\underset{\cdot}{\varrho}\iota}l\underset{\cdot}{\tilde{\imath}}m$	$ku^{\underset{\cdot}{\varrho}\iota}l\underset{\cdot}{\tilde{u}}b^{w}j$	$ku^{\underset{\cdot}{\varrho}\iota}l\tilde{u}ba'$
2. P.	$ku^{\underset{\cdot}{\varrho}\iota}lu.ı$	$ku^{\underset{\cdot}{\varrho}\iota}l\underset{\cdot}{\tilde{u}}.ı^{j}j$	$ku^{\underset{\cdot}{\varrho}\iota}l\tilde{u}la'$
3. P.	$ku^{\underset{\cdot}{\varrho}\iota}luda$, $k\underset{\cdot}{\tilde{u}}^{\lambda\iota}l\underset{\cdot}{\tilde{u}}da$	$ku^{\underset{\cdot}{\varrho}\iota}l\underset{\cdot}{\tilde{u}}d^{j}j$	$ku^{\lambda\iota}l\tilde{u}d\tilde{u}n$
1. P.	$\underset{\cdot}{\dot{z}}\underset{\cdot}{\tilde{o}}^{\iota}l\underset{\cdot}{\tilde{\imath}}m$	$\underset{\cdot}{\dot{z}}\underset{\cdot}{\tilde{o}}^{\iota}l^{\iota}b^{w}j$	$\underset{\cdot}{\dot{z}}\underset{\cdot}{\tilde{o}}^{\iota}l^{\iota}bi'$
3. P.	$\underset{\cdot}{\dot{z}}\underset{\cdot}{\tilde{o}}^{\iota}l\underset{\cdot}{\tilde{e}}t$		$\underset{\cdot}{\dot{z}}\underset{\cdot}{\tilde{o}}^{\iota}l\underset{\cdot}{\tilde{\imath}}^{\iota}j$

Figure 3. Examples of Use of Special Characters in Linguistics Primary and Secondary Literature. Reproduced with permission from A. J. Joki, ed., Kai Donners Kamassisches Worterbüch nebst Sprachproben und Hauptzügen der Grammatik.

REFERENCES

1. Guttsman, W. L. The literature of the social sciences and provision for research in them. Journal of Documentation, 22:3.186-194, 1966.

APPENDIX A

LINCS DOCUMENT CHARACTERISTICS STUDY

I. Bibliographic Entry Elements

The entry elements are presented here as they were used in the study. 44 were defined initially, the remaining seven were added, during the coding. The elements in this list are not grouped by function or location with the entry.

Code	Element
01	Entry number
02	Descriptors
03	Abstract or annotation
04	Author, personal
05	Author, corporate
06	Editor
07	Affiliation without address
08	Affiliation with mailing address
09	Title of work
10	Translation of title
11	Summary statement (summary in other languages)
12	Statement giving conference/congress data
13	Dedication statement (if distinct from title)
14	Biographic statement (if distinct from title)
15	Statement giving purpose or origin of miscellany
16	Other descriptive statement elaborating but not part of title
17	Translation statement (if work is a translation)
18	Translator
19	Original title of work
20	Language or original
21	Statement giving source-bibliography listing of original work (if work was cited in earlier issue of the bibliography)
22	Edition statement
23	Series statement
24	Title of source book
25	Main-entry reference (e.g. in contributions when, instead of full data for the source, a reference to the full source is given)
26	Place of publication
27	Publisher

Code	Element
28	Sponsoring body or agency
29	Year of publication
30	Illustration statement (figures, tables, etc.)
31	Pages
32	Report series and number
33	Contract number
34	Availability (e.g. through document clearinghouses)
35	Order number
36	Price
37	Statement giving form of work (manuscript, unpublished report, reprint, etc.)
38	Title of journal (may represent title of work as in a review of a journal, but generally represents title of source as in a journal article)
39	Journal identification (city and/or country where journal published)
40	Collation (includes volume, issue, part, pagination, year; treated as an entity)
41	Review statement (includes title and collation of reviewing journal; treated as an entity)
42	Reviewer
43	Statement indicating type of work (thesis, dissertation, report, etc.)
44	Degree-granting institution
45	Volume and/or issue (treated as distinct from collation only in reviews of journals)
46	Statement giving publication date of original work
47	Statement on other published form of work (e.g. also available in paper-backed edition)
48	Paper-back series identification
49	Abbreviated distinctive title (if main work is entered elsewhere under full title)
50	Other statement referring to form of work
51	Year degree granted

II. Document Types

Document types are presented here as they were coded in the study. The variations under each type were necessary for the purpose of coding the elements of each but do not functionally distinguish document categories. The codes are entirely arbitrary. Some non-book materials (450, 451, 452) were included in the list but not actually used because, although there are citations for such materials, they are rare.

Code	Type of Publication
	Separately Published Works
	Book
	Minimal elements: author (personal or corporate, or editor), title, place of publication, publisher, year, pages.
100	Book, personal author
101	Book, with edition statement
102	Book, with series statement
103	Book, with edition and series statement
104	Book, with foreign-language title (if differentiated)
105	Book, with foreign-language title and summary in other language(s)
106	Book, with foreign-language title and translation of title
107	Book, editor
111	Book, editor, with series statement
120	Book, corporate author
130	Book, translation of earlier, other-language edition
131	Book, with review
132	Book, with abstract or annotation
	Technical Report
	Minimal elements: personal and/or corporate author, title, report number, contract number, year, originating agency, pages.
150	Report, personal author
151	Report, corporate author
152	Report, personal author and corporate source

Code	Type of Publication
	Thesis or Dissertation
200	Commercially published
	Minimal elements: author, title, place of publication, publisher, year, pages, indication of type, institution.
201	Commercially published, part of series
202	Not commercially published
	Minimal elements: author, title, institution, year, pages, indication of type.
	Congress, Conference, Symposium
250	Published under specific title ("theme") of conference
	Minimal elements: specific title of conference, statement of conference particulars (place, date, number, etc.), editor, place of publication, publisher, year, pages.
251	Published under Proceedings
	Minimal elements: conference particulars as title, editor, place of publication, publisher, year, pages.
300	**Festschrift**
	Minimal elements: title, dedication statement (if not included in title), editor, place of publication, publisher, year, pages.
350	**Biography**
	Minimal elements: title, biographic statement (if not part of title), author or editor, place of publication and publisher (if book), journal source and collation (if article), year, pages.
400	**Miscellany**
	Minimal elements: editor, title, statement of purpose or origin, place of publication, publisher, year, pages.

Code	Type of Publication
	Non-Book Material
450	Audiorecord
	Minimal elements: author, title, statement of form, place of publication, publisher, album and record number, year, pages, diameter of disc, speed of disc, recording channel.
451	Motion picture
452	Filmstrip
	Unpublished Material
500	Minimal elements: author, title, a statement giving purpose or origin (e.g. paper presented at conference), year, pages, statement of form (e.g. manuscript, unpublished report).
	Periodical. Reference to entire volume or issue
600	Review of periodical
	Minimal elements: title of periodical, place of publication, publisher or sponsoring body, volume, issue, year, pages, review (treat review: title of reviewing journal and collation, is an entity), reviewer.

Code	Type of Publication

Articles

700 Article in Periodical

Minimal elements: author, title of article, title of periodical, collation, statement of summary.

Item in a Book

300 Work containing item is not main entry

Minimal elements: author, title of contribution, title of book, editor, place of publication, publisher, year, pages.

801 Work containing item is main entry

Minimal elements: author, title of contribution, title of book, main-entry reference, pages.

APPENDIX B

Examples of Types of Non-Textual Material Considered in Study of Language Sciences Primary Journals

(a)

ᚦᛟᛏᛏᛖᚾ	'of the wolf'	*zuden*	ᛋᛟᛏᛖᚾ	'of charcoal'	*sūden*
ᚦᛗᛖᛚᛏ	'noise'	*melt*	ᛋᚺᛖᛚᛏ	'spangle'	*helä*
ᚦᛗᛁᚳᚳ ᛚ	'sword'	*miekka*	ᛋᛟᛁᛖᚢ ᛚ	'pretty'	*sievä*
ᚦᚢ ᛚ ᛏᛏᛁ	'basin'	*tati*	ᛋᛏ ᛁ ᛚ ᛏᛁ	'aunt'	*täti*
ᚦᚳᛁᛏ ᛚ	'jaws'	*kita*	ᛋᛁᛏ ᛚ	'east'	*itä*
ᚦᛋᛖᛁᚳᚳ ᛚ	'matter'	*seikka*	ᛋᚳᛖᛚᚾ ᛚ	'well'	*selkä*

TABLE 9

(b)

have
N_1 + *have* + N_2:
N_1 + Vt: *have* + N_2 =
A:N_1 Possessive + N_2 =
N_2 + Prep:*of* + N_1

acquire:	*contract/catch:*	*have a relationship with:*
Xperson *has acquired* Xthing ⊃ Xperson *has* Xthing	Xanimal$_1$ *has caught* Xdisease ⊃ Xanimal$_1$ *has* Xdisease	Xperson$_1$ *is* Xperson's Xrel ⊃ Xperson$_1$ *has* Xrel

FIGURE 4

(c)

TONES OF AMOY HOKKIEN

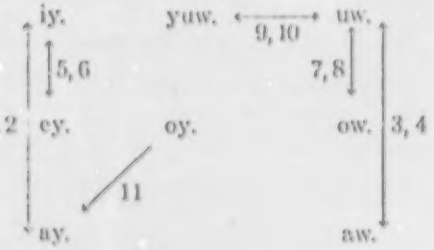
	Ia	Ib	II	IIIa	IIIb	IVa	IVb
	˥	˧˥	˧˩	˩	˧	˩˧	˥
High	+	+	+	−	−	−	+
Falling	−	−	+	+	−	+	−
Rising	−	+	−	−	−	−	−
Long	+	+	+	+	+	−	−

(d)

PENUTIAN KERNEL CORRESPONDENCES

PP	K	M	Ww	Wp	Y	Mi	C
*p	p	p	p	p	p	p	p
*ph	ph		ph	ph	ph		
*k	k	k	k	k	k	k	k
*q	q	k	q	k	x	k	k
*qh	qh				(-k)		
*m	m	m	m	m	m	m	m
*n	n	n	n	n	n	n	n
	w-	w-	w-	w-	w-	w-	w-
	l-	-l-	-l-, -l	-l-, -l	-l-	-l-	-l-, -r-
*r- Initial	s[C, L[V	h	λ, s	λ	ʃh	n	l, r
*-r- Medial	d, l	d	(r?)	r	ʃh	(n?)	r
*-r Final	?	?	r	r	ʃh	n	r
**s)	s-	s-			s-	s-	

(e)



(f)

Atonic		Tonic		
Dissyllabic	Polysyllabic	Dissyllabic	Polysyllabic	
hana-a	*sakura-a*	*hāNa*	*tatémono-a*	(5)
iku-u	*asobu-u*	*kúku-u*	*tabéru-u*	(5)

(g)

Examples of Tables. (Reproduced with the permission of (a, b, e) Language, (c, d) International Journal of American Linguistics, and (f, g) Linguistics.)

(a)

TABLE 1

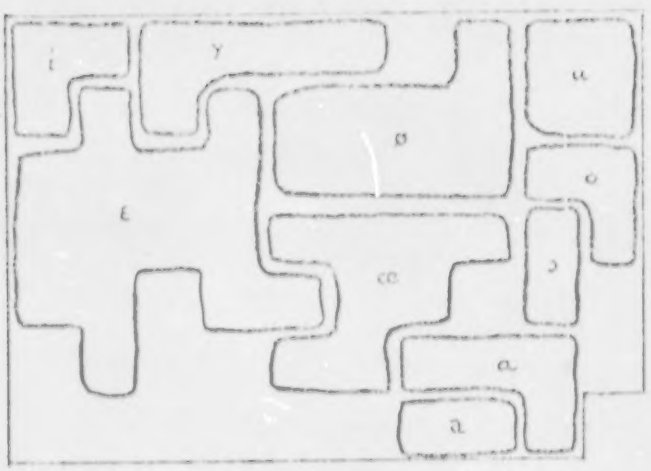
	[illegible]						
	[illegible]		[illegible]		[illegible]		
	-3 (proximate)	-3 (obviative)	-proximate obviative	-4I	-4	-4	-medial
Actor local 1- (speak.)	Refl.	nonoshóhsden Vt-sfs-n Vt, 1211, 1272	[illegible] Vt-s-? Vt, 1211, 1273			nonoshóhsdense Vt-sfs-ne Vt, 1211, 1272, 1281	nonoshóhsdra ~ nonoshóhsdu Vt-s-dr ~ -du Vt, 1211, 1273
2- (addres.)	nonóhsdwdn Vt-s-n Vt, 1212, 1272	Refl.	[illegible] Vt-s-φ Vt, 1214, 1272		[illegible] Vt-di?-fs-n Vt, 1210, 1272		[illegible] Vt-s-dr-φ Vt, 1214, 1271, 1272
zoological prox.- obv.-	nonoshohdíno Vt-sf-no Vt, 1217, 1271	nonoshohdín Vt-sf-n Vt, 1217, 1272	[illegible] Vt-sf-s Vt, 1213, 1273 Refl. nonoshówat Vt-s-t Vt, 1214, 1272	nonoshohdíno? Vt-sf-no? Vt, 1217, 1271, 1274	[illegible] Vt-di?-fs-s Vt, 1216, 1273 [reconstructed]	nonoshohdíno6 Vt-sf-no6 Vt, 1217, 1272, 1281	[illegible] Vt-sf-s? Vt, 1213, 1273 nonoshówat Vt-s-t Vt, 1214, 1273 [by extension of the sg. form]

(b)

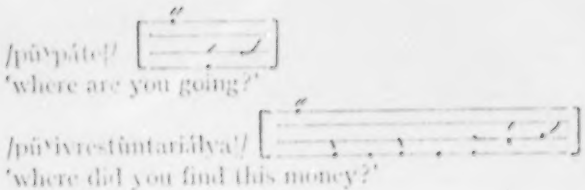
CS:	3.4.78 (see Table 2)	*tip*	*tar*	*fhi*	*sip*	*thas*	*tha*	*mip*	*tas*	*mas*
	Secondary endings (Table 9)	*t*	*tām*	*an*	*s*	*tam*	*ta*	*am*	*ta*	*ma*
	3.4.109-10 (*ghi* → *jas*)	-	-	**us*	-	-	-	-	-	-
	7.3.96 (Incr. If)	**ŧ*	-	-	**ts*	-	-	-	-	-
Char:	3.1.43 (Suffix *sti*)	x	x	x	x	x	x	x	x	x
	3.1.44 (*sic* replaces *sti*)	x	x	x	x	x	x	x	x	x
	7.2.73 (*sis* replaces *sic*)	-	**x*	**x*	-	**x*	**x*	**x*	**x*	**x*
	2.4.77-8 (*ŝ* replaces *sic*)	**x*	**x*	**x*	**x*	**x*	**x*	**x*	**x*	**x*
RESULT:	sibilant aorist	*sīt*	*sītam*	*sunḥ*	*sīs*	*sīstam*	*sīsta*	*sīsam*	*sīsta*	*sīsma*
	sig-aorist	-	*sistām*	*sīsuḥ*	-	*sistam*	*sista*	*sīsam*	*sīsta*	*sīsma*
	root-aorist	*t*	*tām*	*us, an*	*s*	*tam*	*ta*	*am*	*ta*	*ma*

TABLE 14. NON-THEMATIC AORIST

(a)



(b)



Examples of Graphs. (Reproduced with the permission of (a) Language and Speech, and (b) Lingua.)

(a)

Solution I follows below.

1. $\begin{bmatrix} +\text{cns} \\ -\text{cnt} \\ -\text{vd} \end{bmatrix} \rightarrow \begin{bmatrix} +\text{cnt} \\ +\text{vd} \end{bmatrix} / [+\text{voc}] - [+\text{voc}]$

$[+\text{voc}] \rightarrow [+\text{stress}] /$

2. $s \begin{cases} [+\text{cns}] - [+\text{cns}]_{\#}^{\#}[+\text{voc}] \\ [+\text{cns}]_{\#}^{\#}[+\text{voc}][+\text{cns}]_{\#}^{\#}) \end{cases}$

$A \rightarrow [-\text{vd}] /$

3. $\begin{cases} 1) \quad s \\ 2) \begin{bmatrix} +\text{voc} \\ -\text{stress} \end{bmatrix} \begin{bmatrix} +\text{cns} \\ -\text{vd} \end{bmatrix} \begin{bmatrix} +\text{voc} \\ +\text{vd} \end{bmatrix} \\ 3) \begin{bmatrix} +\text{cns} \\ - \end{bmatrix} \begin{bmatrix} +\text{voc} \\ -\text{vd} \end{bmatrix} \end{cases}$

4. $X \rightarrow \begin{cases} [-\text{voc}]/[+\text{voc}] - s \\ \phi/\begin{bmatrix} +\text{voc} \\ -\text{vd} \end{bmatrix} - \begin{bmatrix} +\text{cns} \\ -\text{vd} \end{bmatrix} \end{cases}$

5. $\begin{bmatrix} X \\ (a\,\text{stress}) \end{bmatrix} \begin{bmatrix} X \\ (-a\,\text{stress}) \end{bmatrix} \rightarrow \begin{bmatrix} X \\ +\text{long} \\ +\text{stress} \end{bmatrix}$

(b)

On peut déduire de là que la probabilité d'une voyelle (A) est à la limite:

$$P_\infty(A) = \frac{p(A|B)}{p(B|A) + p(A|B)} = \frac{1}{0{,}49 + 1} = 0{,}671$$

(c)

(1) $\begin{bmatrix} -\text{cons} \\ +\text{vocl} \\ +\text{grav} \\ +\text{comp} \\ +\text{flat} \end{bmatrix} \rightarrow [-\text{flat}] \; / \; \text{——} \begin{bmatrix} +\text{cons} \\ -\text{vocl} \\ +\text{comp} \\ +\text{grav} \end{bmatrix}$

(2) $\begin{bmatrix} -\text{cons} \\ +\text{vocl} \\ +\text{grav} \\ +\text{comp} \\ -\text{flat} \end{bmatrix} \rightarrow [+\text{flat}] \; / \; \text{——} \begin{bmatrix} +\text{cons} \\ -\text{vocl} \\ +\text{comp} \\ +\text{grav} \end{bmatrix}$

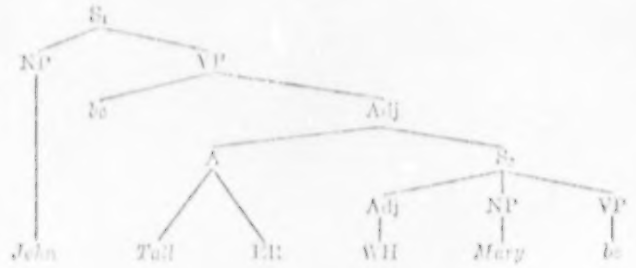
Examples of Mathematical Expressions. (Reproduced with the permission of (a) International Journal of American Linguistics, and (b, c) Linguistics.)

(a)

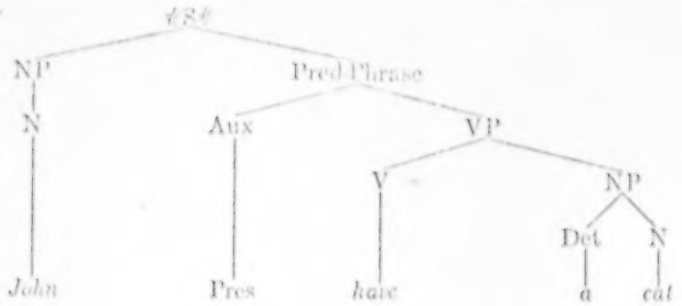
(b)

(c)

FIGURE 2



(d)



Examples of Tree Structures. (Reproduced with the permission of (a, c, d) Language, and (b) Linguistics.)

(a)

	underlying	*phonetic*
друг	(drúg)	drŭg
дружить	(drug + i + t')	družit'
дружу	((drug + i + ú) + ú)	družú
дружат	((drug + i + ú) + at)	družát
веку	((p'ok + o) + ú)	p'ikŭ
печешь	((p'ok + ó) + š)	p'ič'óš
печет	((p'ok + ó) + t)	p'ič'ót

(b)

76. *ki* ~ *-pki* 'four': *qkpĭi* 57, *cĭpĭi* 22, *thăpĭi* 37, *wăpĭi* 38.
77. *-pele* 'thing': *lăpele* 25, *tăpece* 29, *tăpele* 31, *tapeke* 39.
78. *tale* ~ *-ptale* 'to play': *qhăptălĕ* 47, *thăptăle* 7, *lăptsĭă* 30, *tăptăe* 34
79. *kă* ~ *-pkă* 'hundred': *qkpkă* 16, *tapkă* 2, *šipkă* 8.
80. *-pkăa* 'step, base': *qăpkăa* 52, *kăpkăa* 34, *tăpkăe* 12.
81. *kŭq* (*kăpkŏ*) ~ *-pkŏa* 'crooked': *qăpkŏa* 66, *qhapkŏa* 21, *qapkŏa* 42.
82. *ek* ~ *-peă* 'ten': *qăpeă* 19, *ŝăpeă* 2, *šipeă* 8.
83. *ehă* ~ *-peă* 'third': *ehăpeă* 7, *tĭpeă* 27, *lăpeă* 45.
84. *tha* ~ *-ptha* 'mark': *šiptha* 24, *šapkha* 9, *kăpkha* 15.
85. *tlx* ~ *-pčix* 'to say, tell': *papčix* 16, *qapčix* 17.
86. *-pčhat*: *ehăpchă* 7, *tlăpchăe* 60.
87. *pła* ~ *-pčła* 'to bake': *šapčła* 9, *ălpčła* 49.

(c)

18) *nivaq* 'wheat, grain' (Goth. *hwaiteis* 'wheat', *hwits* 'white'), *ŝ-ŋiaŋ* *ŋŏoz* (Laf. *n-dzi-ma*, p.157), *ŋaiaq*, *ŋiŋŋ* 'a sea fish' (E. *carp*, p.162f.), *ŋŏiaŋa* 'celery, parsley' (= *ĝel-* OCS. *zlŭva* 'grass'; Alt. *čŋ*, *θŋŋa* (Gr. χόλγος, p.159f.), etc.; *ŋhl* pointed in ZVI 41); *ŋaŋŋa* (= *exphatus*) and *iaŋŋa* (= *exhitus*) (p. 164); *ŋiŋa* 'tower, castle, fortress' (= *dhrghis*, Skt. *dṛ́ṃhati* 'fasten'), *ŋŋaŋŋaa* 'dry twigs, dry wood' (= *dhrangh-*, ON *drangr* 'upright trunk of a tree', NS *drŭgi* 'dry').

(d)

(e.g., *to lie*, *to drag*, *breast*).

Nasalized vowels: *Yıs > ǫ, ą, į, ǫ (e.g., *same*); *Yın > ą, ą, į, ǫ (e.g., *bread*); *Yɪ > ų, ų, į, ǫ (*to sag*, *to axe*, *to rock*, *to drink*); *Yɜ > ę, ę, ę, ą (e.g., *Yɜ *to come* > *ʔę, *ʔę, *ʔ ą, ą); *Yɐ > ą, ą, ǫ, ǫ (e.g., *ground*); *Yıɐ > į̨, ę, ǫ, ǫ(ɾ) (e.g., *tooth*); *Yąɐ > į̨, ą, ǫ, ǫ (e.g., *Yąɐn *four* > *ʔį̨n-, *ʔą̨n, *ʔę-, ǫ-n-); *Yɛ > ų, ų, ǫ, ǫ (e.g., *ɟǫYɛ *sound* > ɟų-, ɟųn *bell*, -ɟǫ, ɟǫ e).

(e)

/ #ɔb / bʌvt / ðɪyowal / hθ / ðɪyN / nuɑn / ntyɔvəŋ / fað / ðɔw / wɪʃ/ ʃɪz / zɪz / tus / sɪyə / plɔʒ / ʒur / ɪɛn / dʒyoyən / nəf / fes / səv / vɪz / gudɔ / klin / / tʃəyɔidə /

(f)

poème grec de la stèle nous aide a devmer le sens :

ὠδδεῖς πεσ Λυσίον σεῖῤῥε, τοιάνδ' ἀνέθηκεν
δώδεκα θεοῖς ἀγορᾶς ἐν καθαρᾷ τεμένει

Examples of Special Symbols. (Reproduced with the permission of (a, e) Linguistics, (b) Language, (c) Lingua, (d) International Journal of American Linguistics, and (f) Bulletin de la Societe de Linguistique de Paris.)

APPENDIX C

Bibliography of Primary Works in a Sample of Computational-Linguistics Literature

Abernathy, Robert. Mathematical linguistics. In: Thomas A. Sebeok, ed., Current trends in linguistics. I. Soviet and East European linguistics (The Hague: Mouton, 1963).

Akhmanova, O. S.; Melcuk, I. A., Frumkina, R. M., Paduceva, E. V. Exact methods in linguistic research. Transl. by David G. Hays and Dolores V. Mohr. Santa Monica, Calif.: Rand, 1963.

American Mathematical Society. Structure of languages and its mathematical aspects. Symposium in applied mathematics 12. Providence, Rhode Island: American Mathematical Society, 1961.

Bobrow, D. G., Fraser, J. B., Quillian, M. R. Automated language processing. In: Carlos A. Cuadra, ed., Annual review of information science and technology 2 (New York: Interscience Publishers, 1967).

Borko, Harold. Automated language processing. New York: John Wiley, 1967.

Carnap, Rudolf. Foundations of logic and mathematics. In: Jerry A. Fodor and Jerrold J. Katz, eds., The structure of language. Readings in the philosophy of language (Englewood Cliffs, N. J.: Prentice-Hall, 1964).

Carroll, John B. The statistical study of language. In: John B. Carroll, The study of language (Cambridge, Mass.: Harvard University Press, 1953).

Chomsky, Noam. Formal properties of grammars. In: R. Duncan Luce, Robert R. Bush, Eugene Galanter, eds., Handbook of mathematical psychology (New York: John Wiley, 1963).

---. Syntactic structures. The Hague: Mouton, 1962.

Chomsky, Noam; Miller, George A. Introduction to the formal analysis of natural languages. In: R. Duncan Luce, Robert R. Bush, and Eugene Galanter, eds., Handbook of mathematical psychology (New York: John Wiley, 1963).

Delavenay, Emile. An introduction to machine translation. London: Thames & Hudson, 1960.

Edmundson, H. P., ed. Proceedings of the national symposium on machine translation, University of California, Los Angeles, February 2-5, 1960. New York: Prentice-Hall, 1961.

Garvin, Paul L. Natural language and the computer. New York: McGraw-Hill, 1963.

Garvin, Paul L.; Spolsky, Bernard, eds., Computation in linguistics. Indiana University studies in the history and theory of linguistics. Bloomington: Indiana University Press, 1966.

Harper, Kenneth E. Machine translation. In: Thomas A. Sebeok, ed., Current trends in linguistics. I. Soviet and East European linguistics (The Hague: Mouton, 1963).

Hays, David G. Introduction to computational linguistics. New York: American Elsevier, 1967.

-----., ed. Readings in automatic language processing. New York: American Elsevier, 1966.

Herdan, Gustav. The calculus of linguistic observations. The Hague: Mouton, 1962.

-----. Language as choice and chance. Groningen: P. Noordhoff, 1956.

-----. Type-token mathematics: a textbook of mathematical linguistics. S.-Gravenhage: Mouton, 1960.

Hockett, Charles F. Language, mathematics, and linguistics. In: Thomas A. Sebeok, ed., Current trends in linguistics. III. Theoretical foundations (The Hague: Mouton, 1966).

Ivic, Milka. Mathematical linguistics. In: Milka Ivic, Trends in linguistics (The Hague: Mouton, 1965).

Klefer, Ferenc. Mathematical linguistics in Eastern Europe. New York: American Elsevier, 1968.

Kuno, Susumo. Computer analysis of natural languages. In: American Mathematical Society, Proceedings of symposia in applied mathematics 19 (Providence, Rhode Island: 1967).

Lehmann, Winfred P. Computational linguistics. In: Arnold W. Pratt, A. Hood Roberts, and Kathleen Lewis, eds., Seminar on computational linguistics (Bethesda, Maryland: National Institutes of Health, 1965).

Locke, William N., Booth, Donald A., eds. Machine translation of languages. Cambridge, Mass.: Technology Press of MIT, 1955.

National Physical Laboratory. 1961 international conference on machine translation of languages and applied language analysis, 1, 2. London: Her Majesty's Stationery Office, 1962.

Oettinger, Anthony G. Automatic language translation. Harvard monographs in applied science 8. Cambridge, Mass.: Harvard University Press, 1960.

Oettinger, Anthony G. Linguistics and mathematics. In: Ernst Pulgram, ed., Studies presented to Joshua Whatmough (The Hague: Mouton, 1957).

Papp, Ferenc. Mathematical linguistics in the Soviet Union. The Hague: Mouton, 1966.

Plath, Warren. Mathematical linguistics. In: Christine Mohrmann, Alf Sommerfelt, and Joshua Whatmough, eds., Trends in European and American linguistics, 1930-1960 (Utrecht: Spectrum, 1961).

Salton, Gerard. Automated language processing. In: Carlos A. Cuadra, ed., Annual review of information science and technology 3 (Chicago: Encyclopaedia Britannica, 1968).

Simmons, Robert F. Automated language processing. In: Carlos A. Cuadra, ed., Annual review of information science and technology 1 (New York: Interscience Publishers, 1966).

Whatmough, Joshua. Mathematical linguistics. In: Proceedings of the 8th International Congress of Linguistics (Oslo: 1958).

Yngve, Victor H. A model and an hypothesis for language structure. In: Proceedings of the American Philosophical Society 104 (American Philosophical Society, 1960).